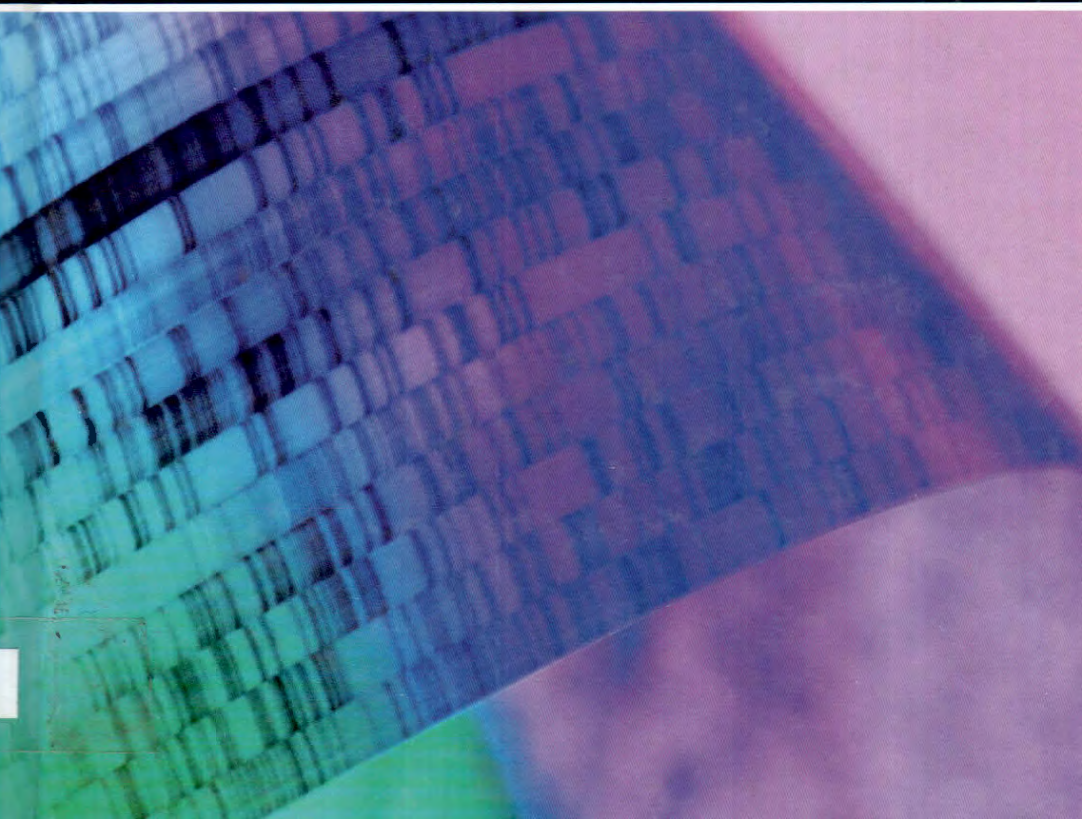
NATIONAL GEOGRAPHIC

EL CÓDIGO GENÉTICO

El secreto de la vida



Si el ADN es la molécula que porta la información biológica de la mayoría de los seres vivos de la Tierra, el código genético es el diccionario, que permite que esa información se traduzca a proteínas, las moléculas encargadas de hacer que todo organismo funcione. Es, por tanto, el que nos posibilita la comprensión del lenguaje en el que está escrita la vida, tan eficiente que apenas ha conocido cambios desde su origen. Descifrarlo fue un reto que involucró a científicos de todo el mundo y que no solo nos ha ayudado a entender mejor el funcionamiento de nuestro organismo, sino también poder corregir enfermedades de origen genético. Tecnologías como CRISPR, que permite editar de forma precisa el ADN, cortando y sustituyendo secuencias dañadas, ofrecen posibilidades revolucionarias en este campo.

Salvador Ferré es biólogo y divulgador científico.
Mireia Ramos es genetista y divulgadora científica.

EL CÓDIGO GENÉTICO

El secreto de la vida

SUMARIO

Introducción 7

01 **Las reglas que rigen la vida** 13

02 **El desciframiento del código genético** 43

03 **Cambios en la información genética: las mutaciones** 75

04 **Más allá del código: la modificación de genes** 103

Lecturas recomendadas 135

Índice 137

INTRODUCCIÓN

Toda forma de vida sobre el planeta Tierra tiene un código vital común, un libro de instrucciones con indicaciones precisas para generar una bacteria, una mariposa o un gran simio. Ese lenguaje es único y tan eficiente que la evolución apenas lo ha modificado desde su origen. Su desciframiento nos ha permitido entender cada vez mejor nuestra historia evolutiva, nuestros patrones de salud y enfermedad e incluso ha abierto vías a la sanación de enfermedades hasta ahora incurables.

Desde hace décadas, la ciencia ha tratado de desvelar toda la información que esconde el ADN, esa larga molécula de ácido desoxirribonucleico que almacena las instrucciones genéticas que permiten el desarrollo y funcionamiento de todos los seres vivos, así como la transmisión de los caracteres hereditarios. Y es que dado que nuestro organismo, nuestras características e incluso nuestras emociones están reguladas por esas instrucciones contenidas en el ADN, la potencialidad que nos da conocer su lenguaje parece casi ilimitada, especialmente en el ámbito de la biomedicina, ya que ofrece no solo la posibilidad de entender las enfermeda-

des hereditarias y muchas otras con base genética, como el cáncer, sino también de llegar a modificar su curso. Corregir las partes de nuestro genoma que no funcionan correctamente solo es posible si, de manera inequívoca, somos capaces de leer, identificar e interpretar las instrucciones del ADN. Y para ello es esencial conocer el código genético.

El conocimiento del código genético ha permitido diseñar la mayoría de las herramientas básicas que acualmente utiliza cualquier laboratorio de genética o biología molecular y, gracias a ese conocimiento, se han desarrollado numerosas técnicas que están abriendo nuevos campos de investigación. La terapia genética es una de ellas, entre cuyos objetivos se encuentra la activación o desactivación de determinados genes para revertir su mal funcionamiento. Actualmente, la técnica que presenta unas posibilidades más revolucionarias es la denominada CRISPR, que utiliza una molécula (Cas) con la que es posible cortar regiones específicas de ADN e incluso sustituirlas por otras. De este modo podemos manipular la información genética, los genes en definitiva, con un nivel de precisión nunca antes alcanzado. Esta técnica se basa en un mecanismo bacteriano de defensa, funcionalmente equivalente a nuestro sistema inmunológico y, dado que la estructura de ADN y el código genético son prácticamente universales, podemos usar dicha técnica en nuestra especie.

Pero ¿cuál es el papel que desempeña el código genético? El lenguaje del ADN está formado por cuatro letras correspondientes a cuatro nucleótidos de desoxirribosa que contienen bases nitrogenadas distintas: adenina (A), guanina (G), citosina (C) y timina (T). Pues bien, el código genético consiste en un sistema que permite la traducción de la información escrita en ese lenguaje de nucleótidos, el lenguaje de los genes, a otro lenguaje que es el de las proteínas, moléculas formadas por uniones de aminoácidos, que intervienen en nuestra formación anatómica y funcionalidad fisiológica. Realmente, el código genético interpreta la información contenida en

nuestros genes dando lugar a los caracteres hereditarios capaces de pasar a las sucesivas generaciones.

Estos caracteres hereditarios ya los había observado en 1866 el monje agustino y gran naturalista austríaco Gregor Mendel cuando investigaba en su jardín los patrones de herencia en las plantas de guisantes. Su minucioso estudio dio lugar a lo que hoy conocemos como las bases de la genética. Pero Mendel no sabía nada del ADN, ese ácido nucleico incluido en los cromosomas que se encuentran en el interior del núcleo de nuestras células, y mucho menos de su modo de información y lectura. Tuvo que pasar casi un siglo, hasta 1953, para que dos investigadores, el estadounidense James Watson y el británico Francis Crick, presentaran la estructura terciaria de doble hélice del ADN, un descubrimiento que iba a representar un antes y un después en nuestro conocimiento sobre la genética. Cinco años más tarde, el propio Crick formuló una hipótesis que señala que el flujo de información genética en los sistemas biológicos va siempre en la misma dirección: el ADN se replica antes de que la célula se divida, con el fin de dotar a cada célula hija de una copia idéntica del material genético; después, una de sus cadenas —llamada cadena codificante— se transcribe a ácido ribonucleico mensajero (ARNm) —una molécula formada por una cadena monocatenaria de ribonucleótidos—, y finalmente esta se traduce a proteínas. A partir de este momento comenzó a surgir la idea de la existencia de un código que fuese capaz de trasladar la información contenida en la secuencia de los nucleótidos del ADN (los cuatro ya mencionados A, G, C y T) hasta la secuencia de los aminoácidos presentes en las proteínas (20 aminoácidos diferentes).

Descubrir este código genético, como se comenzó a llamar desde sus inicios, fue el horizonte de varios centros de investigación a principios de la década de 1960. De hecho, un grupo de científicos organizados en lo que se llamó el Club de las Corbatas de ARN (RNA Tie Club), fundado por el físico y divulgador ucraniano George Gamow, empezó a investigar con una molécula estrechamente

emparentada con el ADN, pero que, a la vez parecía tener un papel decisivo para la producción de proteínas: el ARNm. Poniendo sobre la mesa conocimientos de combinatoria básica, se pudo establecer el lenguaje del ADN. El secreto estaba en leer la secuencia de los desoxinucleótidos del ADN (A, G, C y T) de tres en tres. En el ARNm estos tripletes, llamados codones, presentan 64 posibles combinaciones. Cada uno de esos 64 tripletes se corresponden con alguno de los 20 aminoácidos de las proteínas o con algún triplete de señalización de inicio o de paro.

Así, las células de cualquier organismo interpretan la información del ADN en secuencias de tripletes usando invariablemente el código genético. Una alteración en alguna de las bases del ADN hace que el código interprete un aminoácido erróneo que se colocará en la proteína. Estas variaciones son mutaciones, cambios en la secuencia del ADN, que pueden llegar a tener efectos perjudiciales para nuestra salud. Estas mutaciones pueden deberse a errores a la hora de copiar el ADN, errores propios de las células que pueden perpetuarse en el material genético de las mismas. Pero también existen factores externos que inciden de manera continuada sobre el ADN, como la radiación solar, el humo del tabaco o incluso algunos tipos de virus.

Por lo tanto, conociendo el código genético que ordena la traducción del ADN a proteínas podremos saber qué consecuencias tienen las mutaciones para la célula. Es en lo que se basa el diagnóstico genético que fue la primera aplicación de la genética a la salud humana. La posibilidad de detectar enfermedades incluso antes del nacimiento de la persona o de establecer cuáles son aquellas a las que estamos predispuestos supuso un salto de gigante en la prevención y tratamiento, si bien no de todas, sí de numerosas enfermedades con base genética. Es por ello que en muchos hospitales ya se ofrece un servicio de consejo genético familiar. El trabajo de estas unidades se basa en identificar patrones hereditarios dentro de una familia que puedan ser los causantes de provocar o favorecer la aparición de una determinada dolencia.

El exitoso comienzo de la aplicación de los conocimientos sobre genética en el campo de la medicina, facilitando el diagnóstico y asesoramiento genético, llevaron la investigación a un siguiente paso: el tratamiento de enfermedades. En este aspecto, y gracias a la universalidad del código genético, la ciencia ha logrado introducir genes humanos en células simples, como las bacterias, posibilitando que estas actúen como fábricas de proteínas. Como ejemplo paradigmático está la producción de insulina humana en bacterias *Escherichia coli*, lo que permite la «fabricación» a escala industrial de esta hormona para ser utilizada por millones de personas diabéticas.

Pero quizá la terapia más revolucionaria sea la génica, es decir, la modificación de genes para corregir un fallo en el ADN que en ocasiones causa una enfermedad. En este sentido, la tecnología CRISPR ha supuesto un gran avance y una propuesta de futuro. Mediante el uso de esta nueva herramienta de edición de genes, la ciencia está consiguiendo desarrollar nuevas formas de tratar enfermedades como el cáncer, determinadas distrofias musculares o incluso algunos tipos de ceguera. La técnica ya ha salido del laboratorio y las pruebas con humanos han empezado a llevarse a cabo con resultados que pueden calificarse de esperanzadores.

No obstante, queda aún mucho camino por recorrer, ya que las posibilidades terapéuticas que brinda el descifrado del lenguaje de la vida no parecen tener límites. Algunos laboratorios, por ejemplo, están incluso «retando» las instrucciones del código al introducir, en un símil deportivo, nuevos jugadores en el partido. En este sentido, ya se ha conseguido una molécula de ADN artificial que cuenta con dos bases más que las que ha seleccionado la naturaleza; dos bases artificiales que presentan unas características muy similares a las propias del ADN. Los estudios *in vitro* concluyen que esta nueva molécula es capaz de replicarse, es decir, copiarse a sí misma, condición indispensable para la perpetuación que da sentido a la vida, y que abre la puerta a la fabricación de vida con una secuencia artificial.

Por todo ello hoy podemos afirmar que el descifrado del código genético ha sido un poderoso pilar en el desarrollo de la genética, una disciplina que avanza con paso firme hacia uno de los objetivos soñados por la humanidad: la erradicación de las enfermedades y, con ella, la mejora de las condiciones de vida.

01

LAS REGLAS QUE RIGEN LA VIDA

Toda la vida en nuestro planeta existe gracias a un código genético universal que dicta a nuestras células cómo traducir la información del ADN en las proteínas necesarias para vivir. Su utilización ofrece el potencial tecnológico para prevenir y tratar enfermedades genéticas e incluso construir un nuevo futuro biológico.

La lectura del código secreto más íntimo de la vida, el genético, despertó una emoción semejante a la que debió sentir el egiptólogo francés Jean-François Champollion cuando, tras estudiar la piedra Rosetta, logró descifrar la escritura jeroglífica a principios del siglo XIX. O a la que experimentaron los criptólogos que consiguieron interpretar los códigos que transmitían las máquinas Enigma del ejército alemán durante la Segunda Guerra Mundial, lo que permitía a las tropas alemanas comunicarse de forma ininteligible para sus enemigos. Ambas gestas fueron mucho más que un asombroso desafío intelectual. Si la piedra Rosetta abrió la puerta a la comprensión de la civilización del Antiguo Egipto, descodificar los mensajes de las máquinas Enigma supuso el principio del fin de aquella guerra cruenta. De la misma forma, comprender cómo se traduce la información que encierra nuestro código genético, que sin duda constituye un avance científico espectacular, es ante todo una poderosa fuente de conocimiento de dimensiones aún desconocidas, pero relacionadas con un tema de importancia incontestable: el control de la vida.

El código genético es el diccionario que permite que la información contenida en nuestro ácido desoxirribonucleico (ADN), la molécula que determina todas nuestras características como seres vivos y que se transmite de padres a hijos, se traduzca en proteínas, las moléculas que ejecutan todos los procesos vitales. El desciframiento de este código constituye un hito del pensamiento humano y un punto de inflexión histórico en muchas disciplinas, desde la genética a la biología, pasando por la medicina, la biofísica o la farmacéutica y, aunque ya ha producido cambios en nuestro modo de vida, su potencial augura otros que aún no somos capaces de imaginar.

Desvelar las reglas que rigen el entramado celular de la vida ha sentado las bases para poder leer la información vital de diferentes organismos, pero también para modificarla, e incluso reescribirla a voluntad. Un hito que marcó un antes y un después al respecto fue la culminación de la secuenciación completa del ADN humano, nuestro genoma, en el año 2001, lo que supuso la finalización de uno de los proyectos científicos más ambiciosos de todos los tiempos, el Proyecto Genoma Humano. Desde entonces se han seguido secuenciando otros genomas, de modo que actualmente tenemos a nuestra disposición la información genética de más de 60 000 organismos. Pero no solo ha servido para secuenciar la biodiversidad del presente, sino que también ha permitido un fructífero viaje al pasado, pues el estudio del código genético ha desvelado que todos los seres vivos de la Tierra estamos relacionados de una manera más profunda de lo que podíamos llegar a intuir. Las células de todos los organismos que alberga nuestro planeta, desde la inmensa ballena azul, hasta el más minúsculo parásito, funcionan esencialmente de la misma manera. En el núcleo de las células de todos ellos, el ADN controla el buen funcionamiento de la vida y este se hereda de manera prácticamente inalterada de progenitores a descendientes. El proceso de «traducción» del ADN a proteínas sigue siempre las mismas pautas: las del código genético, cuya condición universal permite que las posibilidades actuales de reescritura de la información genética

sean, en la práctica, casi infinitas. Actualmente, por ejemplo, podemos crear nuevo ADN, recombinar secuencias de ADN de dos organismos distintos, prevenir mutaciones, desarrollar terapias génicas, mejorar las capacidades naturales o incluso crear códigos genéticos artificiales. Aunque la revolución reciente más extraordinaria es, sin duda, la que ofrece la técnica CRISPR, con la que se puede editar el ADN a voluntad con una precisión nunca antes lograda.

Ya hace años que se suceden logros asombrosos en el campo de la genética, como por ejemplo transferir un gen humano a una bacteria para que produzca una proteína útil en medicina. Esta técnica se utilizó en 1982 para obtener bacterias productoras de insulina humana, esencial para los diabéticos, y actualmente sirve para generar toda la insulina que se comercializa. También se ha conseguido curar a personas con enfermedades producidas por errores en un solo gen. En 1990, gracias a incipientes técnicas de ingeniería genética, se logró curar a una niña de cuatro años aquejada de inmunodeficiencia combinada grave, una rara enfermedad que afectaba a su sistema inmunológico y la obligaba a permanecer aislada en un medio estéril para minimizar el riesgo de infecciones. Hoy, esa niña ya es una mujer que lleva una vida completamente normal.

Estos logros son solo las primeras muestras de una revolución en ciernes basada en la intervención directa sobre la información genética de la vida sobre la Tierra. Una revolución que se inició gracias a la inteligencia y tesón de numerosos científicos empeñados en conocer los dos grandes idiomas en que se organiza la vida, el de los ácidos nucleicos (ADN y ARN) y el de las proteínas, así como en descifrar el lenguaje en el que ambos se hablan, el código genético.

EL IDIOMA DEL ADN: NUCLEÓTIDOS ENGARZADOS

Confinada en ese valioso libro de instrucciones que constituye el ADN se halla la información que necesita cada célula para saber

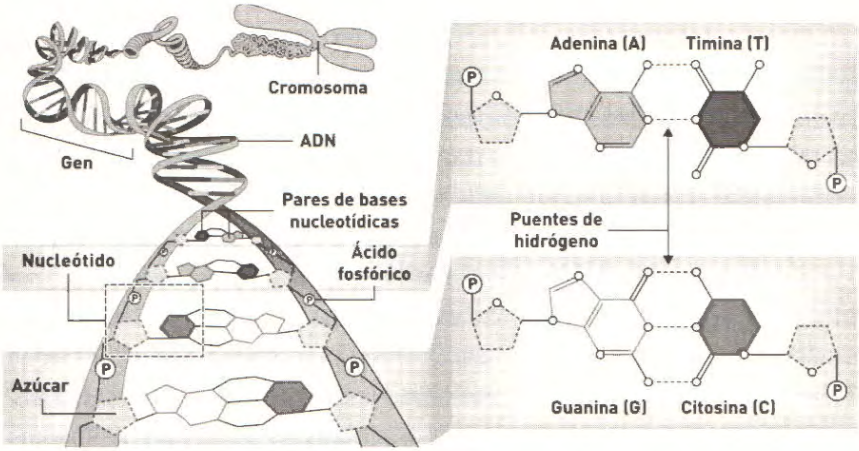
qué proteínas hay que fabricar y cuándo y cómo hay que hacerlo. Pero ¿qué es exactamente el ADN? Químicamente hablando, se trata de una larga molécula formada por la unión secuencial de distintos compuestos, engarzados como si fueran los vagones de un tren, llamados *nucleótidos*, cuya disposición condiciona la información que contiene.

Los nucleótidos, a su vez, son moléculas orgánicas constituidas por una molécula de azúcar, una molécula de ácido fosfórico y una base nitrogenada. Los nucleótidos se unen entre sí mediante su azúcar y su ácido fosfórico, formando una hebra de la que sobresalen las bases nitrogenadas. Cada nucleótido presenta una de las cuatro bases nitrogenadas siguientes: adenina, citosina, guanina y timina y se nombran según la inicial de estas: A, C, G o T. Son las letras del lenguaje genético.

El ADN está formado por dos cadenas que se enrollan formando una doble hélice. Esta, a su vez, se compacta en forma de cromosomas, los cuales se encuentran en el núcleo de las células. La doble hélice está constituida por parejas de nucleótidos complementarias según sus bases nitrogenadas, las cuales se enlazan entre sí siempre de la misma manera: la citosina (C) se une exclusivamente con la guanina (G), mientras que la adenina (A) se une únicamente con la timina (T). Las dos cadenas que forman la doble hélice se unen fuertemente porque una presenta la secuencia «complementaria» de la otra. Las parejas de bases unen una cadena con otra, como si fueran los peldaños de una escalera de caracol a través de puentes (enlaces) de hidrógeno (fig. 1).

La complementariedad de las cadenas permite mantener la secuencia de nucleótidos cuando el ADN se copia, y eso hace que las nuevas células de un organismo hereden siempre la misma información genética. Y es que la herencia se basa en la copia repetitiva de la información de forma fidedigna. Cada una de nuestras células somáticas (no reproductoras) —unos 38 billones— contiene la misma secuencia, la misma información, el genoma individual procedente

Fig. 1



Esquema de la doble hélice del ADN en el que se observa cómo los nucleótidos se disponen en el interior de la cadena y, mediante su complementariedad, mantienen la estructura constante.

de la información del óvulo y del espermatozoide de nuestros progenitores. Cada uno de nosotros tenemos unos 6 000 000 000 de pares de bases alojadas en cada célula somática, ocupando un espacio de unas cinco micras y cada una de estas células es capaz de almacenar aproximadamente unos 1,5 gigabytes de información. Gestionar esta descomunal cantidad de información requiere un alto trabajo de precisión. Así, antes de dividirse, cada célula duplica su ADN. La doble cadena se abre, como una cremallera, y cada mitad sirve de molde a partir del cual se genera su cadena complementaria: de dos cadenas se obtienen cuatro. Una vez que el material ya se ha duplicado, la célula procede a dividirse con la seguridad de que sus células descendientes contendrán la misma información que la que se origina. La complementariedad entre las bases es esencial para que ese libro de la vida pueda ser copiado, teóricamente, sin errores.

EL IDIOMA DE LAS PROTEÍNAS: 20 «LETRAS» ESCRITAS EN «PAPEL ARRUGADO»

En la célula tan importante es el idioma del ADN como el de las proteínas las cuales, recordemos, son las moléculas ejecutoras de las células y responsables directas de casi todos los procesos que tienen lugar en su interior: desde la copia o lectura del propio ADN, hasta el movimiento del citoesqueleto de la célula, pasando por la percepción de estímulos, asegurar la respiración celular o la conexión neuronal. La enorme variedad de procesos en los que intervienen se refleja en la gran diversidad de tamaños y formas de las proteínas. Sin embargo, a pesar de ello, todas y cada una de las proteínas de cualquier organismo vivo están formadas por las mismas 20 moléculas, denominadas *aminoácidos*.

El orden de los aminoácidos que forman la cadena de la proteína determina su función, que puede ser de índole tan distinta como degradar una molécula de azúcar o copiar el ADN. Pero ¿cómo puede una secuencia lineal, donde los mismos grupos químicos se suceden de forma tan diversa, dar lugar a moléculas capaces de realizar funciones tan dispares? Aunque todas las proteínas se sintetizan formando una hilera de aminoácidos unidos entre sí, cada una de ellas se repliega sobre sí misma, adquiriendo una forma particular que determinará su función final. Este plegamiento depende directamente de la interacción fisicoquímica que haya entre los aminoácidos que la forman. Dicho de otra manera, la secuencia lineal de sus aminoácidos establece tanto la configuración como la función de las proteínas. El lenguaje proteico se escribe con 20 letras y sobre papel arrugado.

El cambio de un solo aminoácido por otro puede conllevar la pérdida de la forma característica de la proteína y, por tanto, de su función. Es el caso de la hemoglobina de las personas con anemia falciforme. Esta enfermedad genética deriva de un solo cambio en las 146 letras de una de las proteínas que forma la hemoglobina, esen-

cial para el funcionamiento de nuestro organismo. La hemoglobina se encuentra, principalmente, en las células sanguíneas encargadas del transporte de oxígeno por la sangre, los glóbulos rojos. Si cambia esa letra, los glóbulos rojos de estas personas adquieren una forma diferente de la habitual y no serán tan eficientes a la hora de transportar el oxígeno. ¿La consecuencia? Esas personas sufrirán anemia falciforme, y todo a causa de un cambio aparentemente menor de un solo aminoácido.

> Si la vida tiene un santuario, se encuentra en el núcleo, donde está contenido nuestro ADN, la escritura antigua de nuestro código genético. Y está escrita en un lenguaje que toda la vida puede leer.
>
> NEIL DEGRASSE TYSON

La secuencia de los aminoácidos de una proteína, estrictamente hablando, contiene información, que condiciona la forma de la misma y su función. Así, en el caso de la anemia falciforme, al cambiar la información de la proteína, se modifica también su plegamiento. Salvo en casos muy excepcionales, la información del plegado de las proteínas no se transmite: la información de los aminoácidos queda restringida a la proteína de la que forman parte. Entonces, si la información de los aminoácidos no se transmite, ¿cómo se construye una proteína? ¿Qué determina en qué orden se unirán los aminoácidos en las mismas? Necesitamos a los genes, los fragmentos de ADN que contienen la información necesaria para fabricar las proteínas.

Por su parte, la información que se encuentra en la secuencia de desoxinucleótidos del ADN no determina la forma ni la función del propio ADN, pero sí permite la transformación de la información mediante su copia exacta. El ADN, por sí mismo, no es capaz de ejecutar la información que contiene, es una secuencia a la espera de un sistema de lectura. Y para eso tiene a sus moléculas socias: las proteínas.

Ambas moléculas están, pues, íntimamente relacionadas. En cada núcleo celular, el ADN alberga la información para fabricar

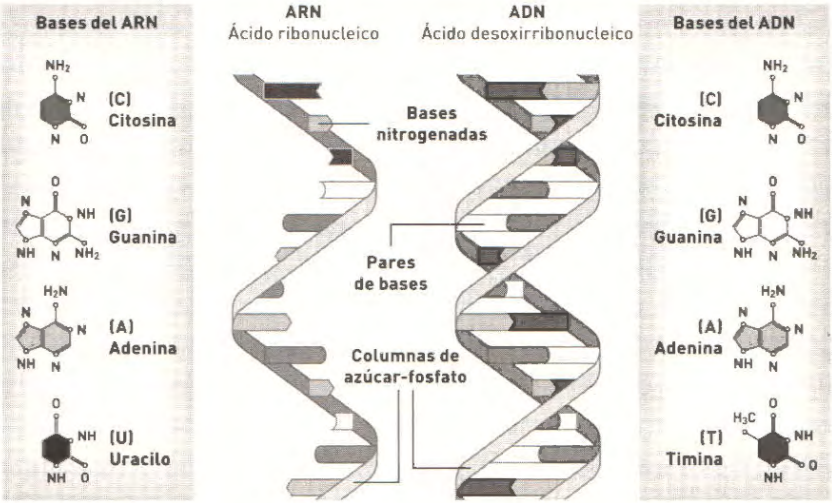
proteínas, un proceso que depende de la acción de otras proteínas que se hallan fuera del núcleo, generadas a partir de la información del propio ADN. Un círculo cerrado que funciona al ritmo del código genético, que es el que posibilita la síntesis proteica. Pero, si el ADN se encuentra en el núcleo de las células y las proteínas se fabrican fuera de él, ¿cómo se establece esa relación? Gracias a las diferentes moléculas de ácido ribonucleico (ARN).

LA TRANSCRIPCIÓN, EL CAMINO HACIA LA SÍNTESIS DE PROTEÍNAS

El ARN es una molécula químicamente muy parecida al ADN. La diferencia entre ambas estriba en la presencia o ausencia de un solo átomo de oxígeno en sus nucleótidos (ahora denominados ribonucleótidos por contener el azúcar ribosa) y en el cambio de una de las letras de su alfabeto. Así, el nucleótido de timina (T) del ADN es en el ARN un nucleótido de uracilo (U). Las moléculas del ARN, por tanto, están escritas con las letras A, C, G y U, en lugar de A, C, G, T. Esa diferencia es sutil solo en apariencia, porque ambas moléculas presentan comportamientos y funciones completamente distintas. El ADN es el guardián de la información hereditaria y, como tal, es estable, se almacena en el núcleo y se escribe en enormes párrafos de centenares de millones de letras (nucleótidos). Por su parte, el ARN tiene una vida más breve, se localiza principalmente fuera del núcleo y se expresa en frases mucho más cortas, de centenares o miles de letras. También su disposición es dispar: mientras que el ADN se presenta en forma de doble cadena, las moléculas de ARN están constituidas por una sola cadena de nucleótidos (fig. 2).

Se han identificado diferentes tipos de ARN, uno de los cuales, el ARN mensajero (ARNm) actúa como intermediario entre la información del ADN y las proteínas; se trata de una copia efímera que permite fabricar proteínas preservando la información genética del

Fig. 2



Representación de las similitudes y las diferencias en cuanto a estructura del ARN y el ADN. Destacan, principalmente, el cambio del azúcar, desoxirribosa a ribosa, y de la base nitrogenada, T a U y la doble hélice característica del ADN.

ADN intacta. El ARN se genera a partir del ADN mediante el proceso que en genética se conoce como *transcripción*, ya que la información no cambia de idioma, sino de soporte: sigue escrito en un alfabeto de cuatro letras, aunque en una molécula ligeramente distinta. Pasa de ADN a ARNm.

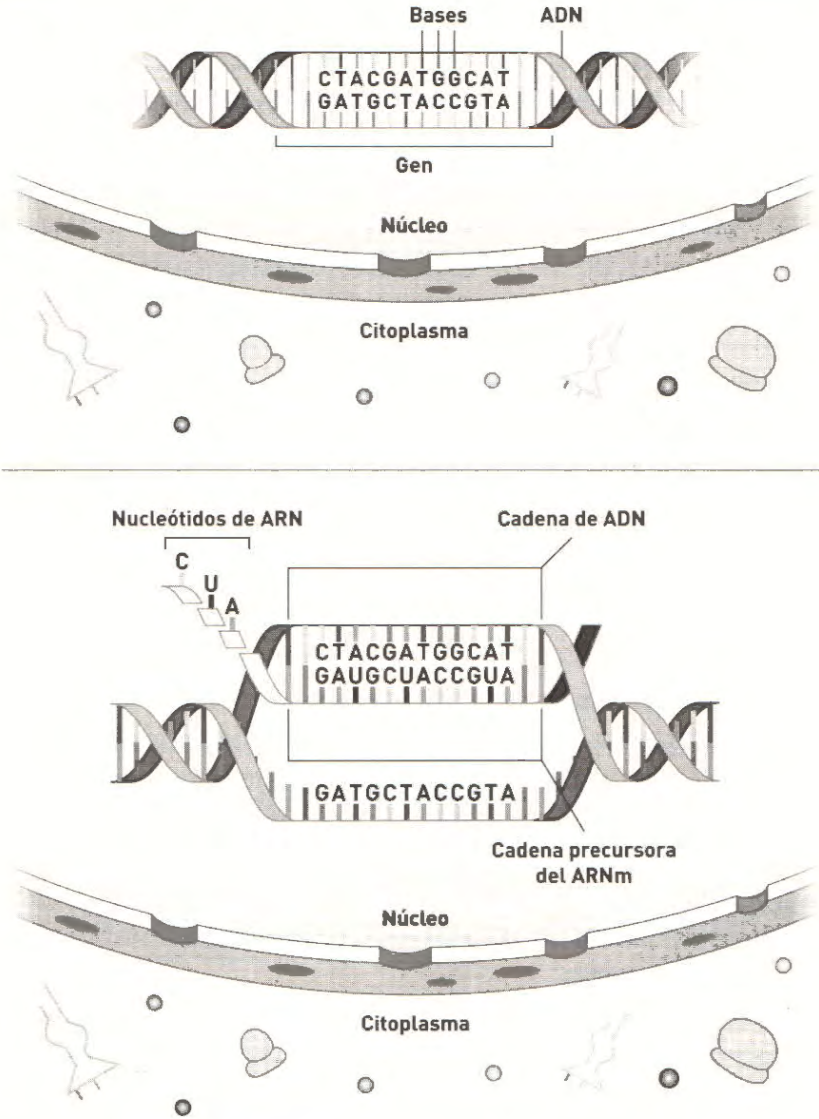
Este proceso funciona así: primero se abre y se desenrolla una pequeña zona de la doble hélice del ADN, y quedan al descubierto las bases de una de las dos hebras del ADN. Es decir, la «cremallera», la escalera de caracol, se abre en dos y a partir de una de las hebras se genera una molécula de ARNm que contendrá una copia de esa información genética. Ese proceso es posible gracias a la acción combinada de diferentes proteínas, aunque la que se ocupa directamente de la producción de la molécula de ARN es una enzima denominada *ARN polimerasa*. Su nombre indica que polimeriza ARN, es decir, que

desencadena un proceso químico —la polimerización— mediante el cual ciertos compuestos, en este caso nucleótidos, se unen de una manera ordenada para dar lugar a una molécula. Lo que hace esta ARN polimerasa es utilizar como molde una de las dos cadenas del ADN y, a partir de esta, sintetiza una molécula en la que va añadiendo nucleótidos copiados y escritos en «soporte ARN», basándose en la complementariedad de las bases. Así, por ejemplo, si lee en el ADN la secuencia AAGG, en el ARN que está sintetizando añadirá la secuencia UUCC, puesto que el uracilo (U) es complementario de la adenina (A) y la citosina (C) es complementaria de la guanina (G). De la misma forma que la complementariedad entre las bases es esencial para la copia del ADN, también lo es para su transcripción fidedigna a ARNm (fig. 3). La información genética se copia mediante la complementariedad entre las bases que forman tanto el ADN como el ARN.

Modificaciones en el ARNm antes de salir del núcleo

Una vez realizado el proceso de transcripción, el ARNm ya está casi listo para viajar hasta el citoplasma, la parte de la célula que rodea el núcleo y que está limitada por la membrana exterior. Pero antes de abandonar el núcleo para ser finalmente traducido, debe someterse a diferentes modificaciones: una de ellas consiste en la eliminación de algunas secuencias llamadas *intrones* que no codifican para proteínas, para permitir que se unan entre sí las secuencias que sí son codificantes, los *exones*. Eso se conoce como *procesamiento por corte y empalme* (*splicing*, en inglés). Una molécula de ARNm puede tener hasta 70 intrones, los cuales pueden presentar tamaños de entre 80 y 10 000 nucleótidos. Cuando estos se eliminan, se mantiene el orden de los exones de forma correlativa, y se pasa de tener una frase sin sentido biológico a otra con significado. Un ejemplo: imaginemos un gen cuya información está escrita en cuatro fragmentos o exones (*a*, *b*, *c* y *d*), separados por secuencias sin sentido o intrones, que

Fig. 3



Arriba, doble cadena de ADN; abajo, el proceso de transcripción, el ADN se abre y se copia la información de una de sus cadenas a ARNm por complementariedad de las bases.

representaremos con una raya (—). En una primera fase tras la transcripción, el ARNm inmaduro contiene tanto exones como intrones: *a—b—c—d*. Cuando los intrones (—) se eliminan, los exones se unen y generan ARNm maduro: *abcd*. Si un ARN inmaduro contuviera un mensaje tipo: «Aúnvoysfbpvzbttuegmudinunauvehddqiñadelpie», tras desprenderse de los intrones (subrayados), habría un ARN maduro de tipo: «Aúnvoysinunauñadelpie».

Pero este proceso se puede complicar porque, según cómo se hagan las uniones durante estas modificaciones, lo que se conoce como *empalme alternativo*, el ARNm madurará de una u otra forma, lo que posibilita que, a partir de un mismo gen, se pueden originar secuencias diferentes de ARNm que producirán proteínas diferentes, ganando por lo tanto diversidad.

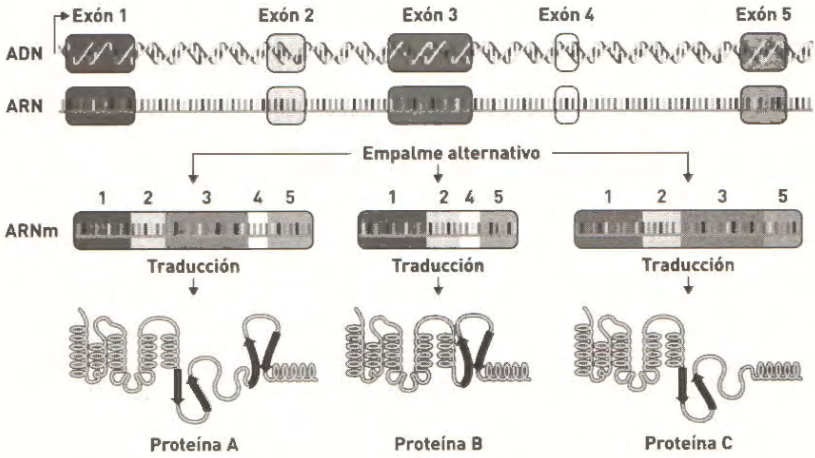
La presencia de empalmes alternativos explica la discrepancia entre la estimación del número de genes humanos y los genes que se identificaron cuando se secuenció nuestro genoma; se estimaba que deberíamos presentar unos 100 000 genes, pero el resultado final arrojó una cifra mucho menor, de alrededor de 25 000. Y es que se creía que cada gen sintetizaba para una sola proteína, cuando hoy sabemos que no: un mismo gen puede sintetizar proteínas distintas. A veces, muchísimas: el gen Dscam de la mosca del vinagre presenta 95 exones, que, por empalme alternativo, generan más de 38 000 ARNm.

El descubrimiento de los intrones, realizado por Phillip Allen Sharp y Richard J. Roberts en 1978, fue muy importante porque estableció de forma definitiva la relación estrictamente unívoca entre un gen y una proteína.

Además del corte y unión de los exones y, por tanto, de la eliminación de los intrones, el ARNm sufre otras modificaciones que tienen como objetivo asegurar que no sea degradado y que se exporte desde el núcleo de la célula al citoplasma. Estas modificaciones incluyen alteraciones de sus extremos. En uno de ellos se añade una base G (guanina) modificada que estabiliza al ARNm y lo protege de proteínas que podrían degradarlo. Esta modificación se conoce como *capping*,

> EL CONTROL Y EL DESCONTROL DEL EMPALME ALTERNATIVO

El empalme alternativo es un mecanismo ubicuo en todas las células eucariotas, es decir, las que cuentan con núcleo, como las nuestras. La capacidad de producir diferentes proteínas (A, B, C en el esquema) a partir de un mismo gen con diferentes exones (el gen del esquema presenta 5 exones) condiciona una diversidad adicional que permite refinar las respuestas de la célula y que implica un control extraordinario. El descontrol del empalme alternativo se ha relacionado con diferentes alteraciones y enfermedades como algunos tipos de cáncer o la diabetes, así como con la senescencia y el envejecimiento. En este sentido, algunas proteínas procedentes de empalmes alternativos minoritarios durante la juventud aparecen con mayor frecuencia en tejidos envejecidos, como los vasos sanguíneos de personas mayores o en las neuronas de personas con alzhéimer.



— En el empalme alternativo, un ARNm inmaduro puede engarzarse de diversas formas, lo que da como resultado ARNm maduros ligeramente distintos que se traducirán a diferentes proteínas.

término que se podría traducir como «formación de casquete o de corona». En el otro extremo se añade una serie de adeninas que puede tener una longitud de hasta 200 nucleótidos. Esta modificación se conoce como *cola de poli-A* y, además de proteger al ARNm de la degradación, facilita su salida del núcleo al citoplasma por los poros nucleares.

Una vez superadas todas esas fases, el ARNm llegará al citoplasma y dará inicio a la traducción de la secuencia de los nucleótidos a la de aminoácidos de una proteína. Es decir, el «idioma ARN» se traducirá en el «idioma proteína».

LA TRADUCCIÓN ENTRE IDIOMAS

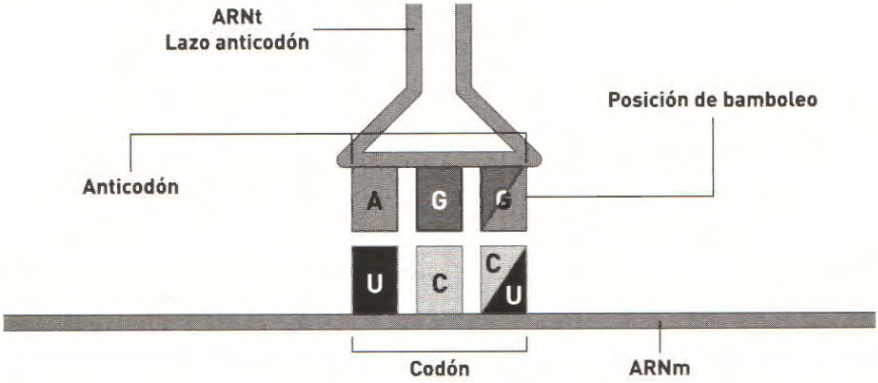
Cuando el ARNm llega al citoplasma, se ponen en marcha las estructuras celulares especializadas en la traducción de la información contenida en el mismo para sintetizar las proteínas. Los ribosomas son una de esas estructuras macromoleculares formados por unas ochenta proteínas y cuatro ARN ribosómicos (específicos para la formación de ribosomas) que se componen de dos subunidades: una subunidad menor y una mayor, que se unen para sintetizar las proteínas y se separan al acabar la síntesis.

La subunidad menor «lee» de tres en tres las secuencias de los ARNm y cuando localiza el triplete (o codón) «AUG», recluta a la subunidad mayor, conformando el ribosoma en el que tendrá lugar la traducción y producción de las proteínas. Los ribosomas avanzan por el ARNm leyendo secuencialmente los codones o tripletes de nucleótidos que constituyen la unidad mínima en que se codifica la información genética.

La producción de una proteína determinada resulta de la unión secuencial de los aminoácidos que transportan los ARN de transferencia o ARNt, moléculas de ARN especializadas en vincular cada codón del ARNm con el correspondiente aminoácido de la proteína a sintetizar. Hay aproximadamente 40 ARNt diferentes en cada

> LA HIPÓTESIS DEL BALANCEO: ¿POR QUÉ HAY MÁS CODONES QUE ARNt?

Los aminoácidos que forman las proteínas no interaccionan directamente con el ARNm, sino que lo hacen mediante el ARNt, el cual presenta en su estructura, por un lado, una serie de tres nucleótidos, los anticodones, que pueden unirse a uno o varios codones del ARNm y, por el otro, un aminoácido específico de los codones que reconoce. El apareamiento entre codón y anticodón se produce siguiendo las normas de complementariedad tanto en la primera como en la segunda base del codón. Sin embargo, en la unión de la tercera base es posible un apareamiento más permisivo. En el ejemplo, A y G del anticodón reconocen claramente a U y C del codón; sin embargo, la G del anticodón puede reconocer tanto a U como a C. Este ARNt, en concreto, se unirá tanto a UCU como a UCC. La hipótesis del balanceo fue formulada antes de disponer de la secuencia de los ARNt. Los resultados experimentales han corraborado la hipótesis: existen ARNt capaces de reconocer diferentes codones, por el balanceo de la tercera base y esta característica es la causante de que haya codones «sinónimos».

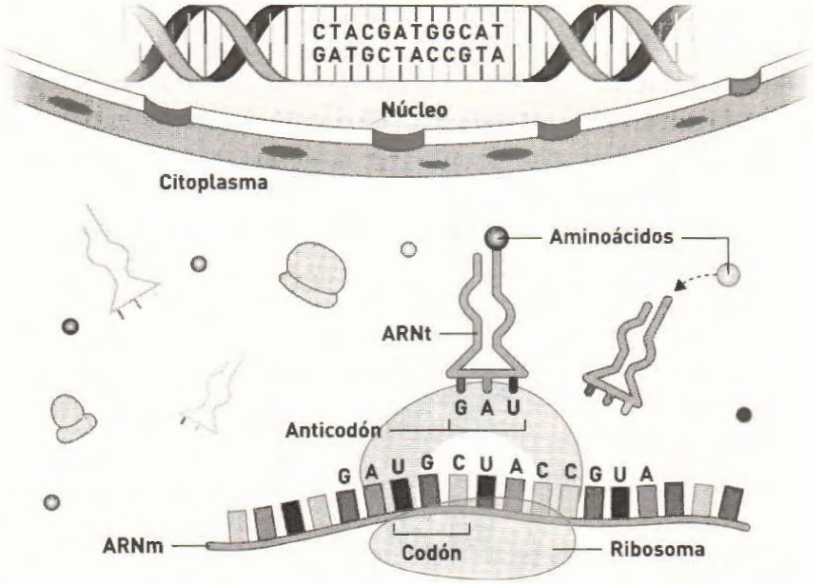


— Esquema del reconocimiento entre el anticodón del ARNt y el codón del ARNm. Se aprecia cómo la tercera letra del anticodón puede reconocer dos letras distintas del codón.

célula y 20 aminoácidos distintos y, por tanto, hay más de un ARNt asociado a un mismo aminoácido.

Cuando un ribosoma lee un trío de palabras (un triplete o codón), se detiene momentáneamente hasta que el ARNt adecuado se une al mismo y su aminoácido se incorpora a la proteína que se está sintetizando. El ARNt es capaz de reconocer el triplete gracias, de nuevo, a la complementariedad entre las diferentes bases del ARN. Así, si el ribosoma se detiene en el codón CUC del ARNm, se esperará hasta que se le una un ARNt que contenga la secuencia complementaria GAG. La complementariedad entre las bases es esencial no solo para la copia del ADN, sino también para la lectura y la traducción de la información genética (fig. 4).

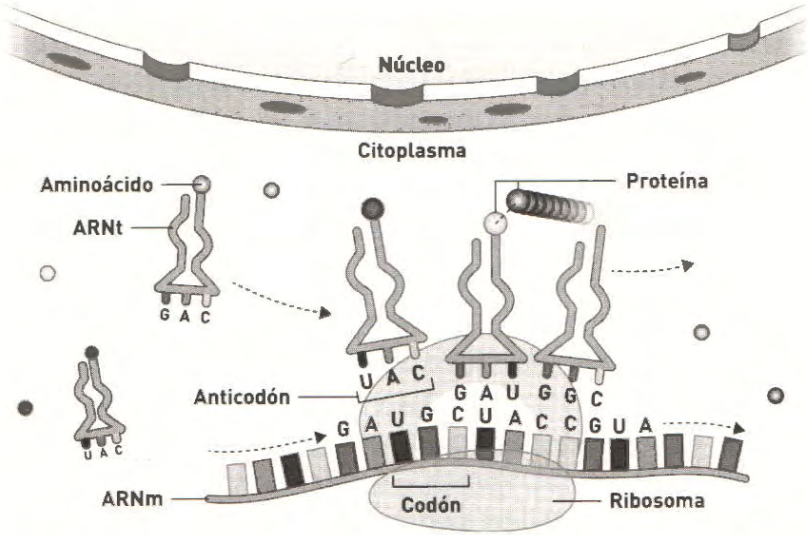
Fig. 4



Esquema de la traducción: el ARNm es leído por los ribosomas. En ellos, los ARNt reconocen los codones específicos y los aminoácidos transportados por los ARNt se unen secuencialmente para formar la proteína.

Cuando se produce esta unión el ARNt «vacío» se desprende del ribosoma y este avanza tres a los siguientes tres nucleótidos del ARNm para leer un nuevo triplete: cada avance se corresponde con un nuevo aminoácido en la proteína. La síntesis continúa hasta que el ribosoma encuentra un codón de finalización, o *codón stop*, una secuencia de tres nucleótidos (UAG, UAA o UGA), que indica al ribosoma que debe detener la traducción. Estos tres codones no son reconocidos por ningún ARNt, sino por proteínas denominadas *factores de liberación*, que contribuyen a la finalización de la síntesis de la proteína y a su liberación para que esta pueda ejercer su función. El ARNm se separa entonces del ribosoma, el cual vuelve a separarse en sus dos subunidades, quedando listas para iniciar una nueva traducción (fig. 5).

Fig. 5



En el ribosoma se sigue sintetizando la proteína a partir del reconocimiento de tripletes del ARNm por parte del ARNt.

En realidad, todo forma parte de un proceso unidireccional basado en la replicación, transcripción y traducción, los tres pasos básicos que llevan a cabo los dos ácidos nucleicos, el ADN y el ARN, y las proteínas, en un planteamiento teórico que debemos fundamentalmente al biólogo británico Francis Crick, y que se conoce como *dogma central de la biología*. En todo sistema biológico, dice la teoría, el flujo de información va siempre en la misma dirección: desde el ADN (en el núcleo) a la proteína (en el citoplasma), pasando por un indispensable intermediario, el ARN. El primer paso se da cuando el ADN se replica antes de que la célula se divida, para dotar a cada célula hija de una copia idéntica del material genético. En segundo lugar, la doble cadena del ADN se abre en dos para que su información pueda ser transcrita en el ARNm y, por último, este se traduce a proteínas. Ese ciclo es unidireccional; sin embargo, se ha observado que los virus pueden replicar su ARN y que este puede realizar una transcripción en sentido contrario para generar ADN de nuevo. A pesar de unas pocas excepciones, en general hoy la formulación del dogma sigue considerándose válida. Se trata, pues, de todo un complejo proceso dirigido por un conjunto de leyes biológicas recogidas en el más universal de los códigos: el genético.

LAS LEYES DEL CÓDIGO GENÉTICO

El código genético es el diccionario que rige la traducción de la secuencia de nucleótidos de un gen a la secuencia de aminoácidos de una proteína. Como se ha comentado, el lenguaje de los ácidos nucleicos se lee de tres en tres nucleótidos, y sus palabras son, siempre, de 3 letras. «Hoy hay col del día» o «Aún voy sin una uña del pie» son frases que podrían pertenecer al idioma de los ácidos nucleicos si no fuera por el exceso de grafías. Cabe recordar que este alfabeto solo tiene 4 letras: A, C, G y U, las correspondientes a las cuatro bases nitrogenadas del ARNm.

Dado que los codones son siempre tripletes de bases y que el alfabeto solo dispone de esas 4 letras distintas para llenar cada una de las tres posiciones, se pueden formar únicamente 64 tripletes o codones, debido a que hay 64 combinaciones posibles de cuatro nucleótidos agrupados en grupos de tres (4^3). El código genético relaciona estas 64 palabras con las 20 letras de las proteínas, los veinte aminoácidos, estableciendo concordancias directas entre ellas. Cada palabra codifica para un solo aminoácido. Resumiendo: el código genético se lee por tripletes y no es ambiguo.

> Así como la palabra mágica de la genética de principios del siglo XX había sido mensaje, se puede decir que la palabra mágica de finales del siglo XX ha sido código.
>
> SIDDHARTHA MUKHERJEE

Al existir más tripletes que aminoácidos (recordemos que varios tripletes se relacionan con el mismo aminoácido) el código genético es redundante. Estos tripletes son sinónimos entre sí. Así, por ejemplo, se pueden encontrar dos tripletes sinónimos, como UGU, UGC, que codifican para el aminoácido cisteína, o AAA y AAG, que codifican para la lisina. Pero lo más habitual es encontrar cuatro tripletes sinónimos para el mismo aminoácido (como GCU, GCC, GCA y GCG para la alanina). Existen tres aminoácidos (leucina, serina y arginina) codificados por seis codones sinónimos (como UUA, UUG, CUU, CUC y CUA, CUG para la leucina). Solamente dos aminoácidos están codificados por tripletes únicos sin sinónimos (AUG para la metionina y UGG para el triptófano).

Pero no todos los tripletes del código se corresponden con un aminoácido, tres no codifican para ninguno y su función es indicar cuándo debe acabarse la síntesis de la proteína, los codones stop. Si ellos indican el final de la transcripción, ¿quién se encarga de señalar cuándo debe empezarse a traducir un ARNm? El codón AUG, que es el que hace que, durante la lectura de tripletes, la subunidad menor del ribosoma «llame» a la subunidad mayor para iniciar la

transcripción. Este triplete AUG se conoce como *codón de inicio* y tiene más de un significado, pues, además de dar el pistoletazo de salida al proceso de traducción, codifica para el aminoácido metionina (Met o M). Esto implica que cualquier proteína se empieza a sintetizar a partir de la metionina, aunque luego, por procesos distintos de la traducción, las proteínas pueden modificarse y perder este primer aminoácido. Si un ribosoma se encuentra con la secuencia AUG una vez que ya se ha iniciado la traducción, añadirá una metionina a la proteína y seguirá adelante, sin que se altere la síntesis de la proteína que está en marcha.

Los tripletes del ARN se leen de manera continua; las palabras del ARN están pegadas unas a las otras, sin nucleótidos que indiquen espacios o comas. Las frases antes utilizadas se escribirían sin separación alguna, como: «Hoyhaycoldeldía» o «Aúnvoysinunauñadelpie». Este mecanismo debe ser muy preciso y sin fallos, pues con los mismos elementos podemos establecer moléculas totalmente diferentes. Si estas ocho letras «a r t o í n e p» fuesen aminoácidos se podrían ordenar para formar la palabra «entropía» que tiene significado, pero también podría hacerse formando la palabra «proteína» que tiene otro diferente.

Otra característica importante es que su lectura es no solapante: tras leer un triplete, se pasa al inmediatamente posterior, con el que no comparte ningún nucleótido. Es decir, la secuencia AUGGCCCUGUGG se lee como AUG GCC CUG UGG; cada letra pertenece únicamente a una palabra. Finalmente, el ARN siempre se lee en el mismo sentido; la frase «Hoyhaycoldeldía» no puede leerse con sentido como «aídledlocyahyoH».

EXCEPCIONES AL CÓDIGO

Aunque el código genético canónico quedó establecido en el año 1966, posteriores estudios e investigaciones condujeron a nuevos

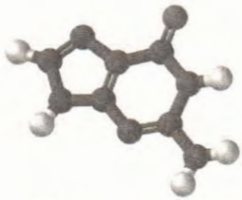
descubrimientos que aclararon algunas de las características del código, matizaron otras e introdujeron nuevas variables a la transmisión de información entre el ADN y las proteínas.

Como ya se ha dicho, el código genético es universal, común para todos los seres de la Tierra, pero también en este caso se cumple aquello de que «la excepción confirma la regla». Ciertamente, la conversión entre ARN y proteínas es exactamente igual en la inmensa mayoría de seres vivos analizados, pero en distintos genomas se han observado cambios en aminoácidos codificados por un determinado codón. La primera de estas singularidades se encontró en un tipo de ADN presente en todas nuestras células: el ADN de las mitocondrias. Estos orgánulos vitales para la célula se encargan de obtener energía a partir del alimento en un proceso durante el cual se consume oxígeno y se libera dióxido de carbono, la respiración celular. Hay mitocondrias en todas las células con núcleo, es decir, en todas las de los animales, plantas, algas, hongos y muchos organismos unicelulares.

Una de sus mayores peculiaridades es que presentan su propio ADN, lo que hace que, dentro de cada una de nuestras células, coexistan dos ADN distintos: el de las mitocondrias, con información relativa a algunas de sus funciones, y el del núcleo, donde se guardan las instrucciones referentes al resto de las funciones de la célula, incluyendo alguna de las mitocondrias. Diferentes evidencias llevaron a postular el origen simbionte (de *simbiosis*, interacción biológica beneficiosa para ambas partes) de estos orgánulos, una teoría que popularizó la bióloga estadounidense Lynn Margulis y que afirma que las mitocondrias procederían de bacterias libres que fueron engullidas por una célula con núcleo. Sin embargo, en vez de ser digeridas, establecieron una relación de mutuo beneficio con la célula que las devoró. Esas bacterias, futuras mitocondrias, encontraron acomodo y protección dentro de la célula con núcleo, mientras que estas, a cambio, obtenían energía de la respiración celular bacteriana. Ambas partes salían

> UN CÓDIGO VITAL

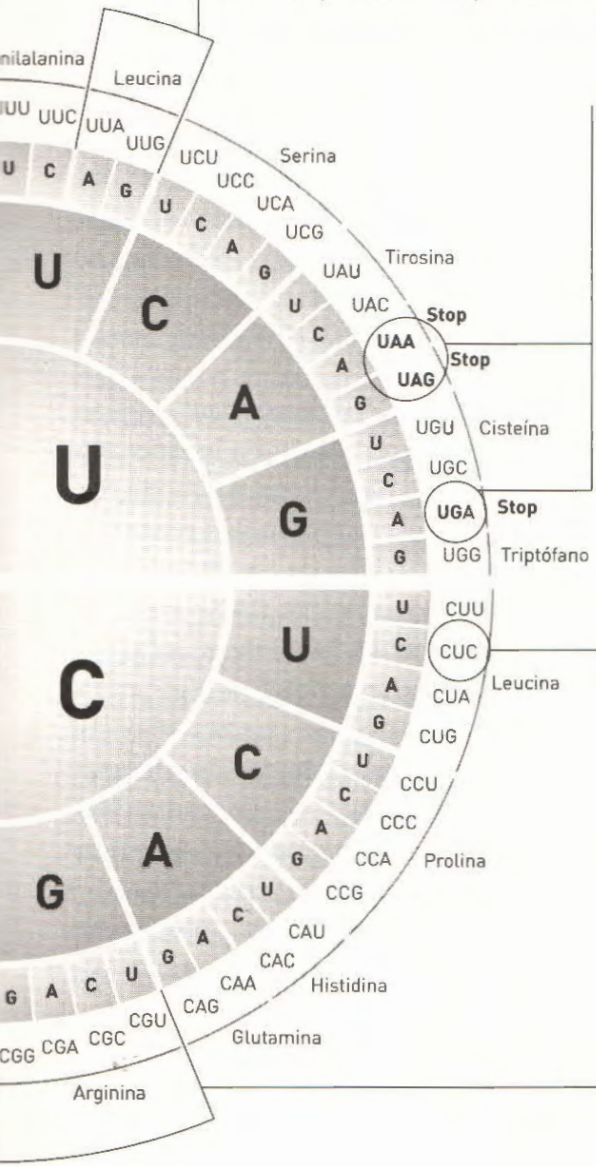
Desde la bacteria más simple, toda la vida en la Tierra se basa en un mismo código genético, cuya representación puede verse en esta tabla circular. En ella, los diferentes tripletes se leen desde el centro del círculo hacia el exterior, es decir, la primera letra del codón se encuentra en el anillo central; la segunda letra, en el anillo intermedio, y la tercera letra, en el anillo exterior. Lo que se indica en este es el codón que se ha construido y el aminioácido para el que codifica dicho triplete.



Las letras mayúsculas indican las bases de los nucleótidos de ARN de los tripletes que se traducirán a proteínas. La figura representa la estructura química de la guanina (G).

El codón AUG codifica para metionina, pero también actúa como codón de inicio. Cuando un ribosoma reconoce este codón en un ARNm, se ensambla definitivamente la maquinaria de traducción y se inicia la fabricación de una nueva proteína.

De los 64 posibles codones, 61 codifican para los 20 aminoácidos que forman las proteínas. En este caso, UUA y UUG codifican para la leucina.



Los codones UAA, UAG y UGA actúan como codones stop. Cuando un ribosoma lee uno de estos codones, se detiene la traducción y la proteína naciente finaliza.

1 codón = 1 aminoácido

El código genético se lee agrupando las letras del ARN (las bases de sus nucleótidos) de tres en tres. Cada palabra de tres letras se conoce como triplete o codón.

El código es redundante: codones distintos codifican para un mismo aminoácido, como en el caso de la arginina, codificada por CGG, CGA, CGC, CGU, AGG y AGA.

ganando; un *win-win*, como se suele decir en la jerga empresarial. Con el paso de los milenios, esta relación se hizo tan estrecha que algunos genes mitocondriales acabaron integrados en el ADN del núcleo, mientras que otros muchos se mantuvieron en el ADN de la propia mitocondria. Hoy sabemos que esta teoría es completamente cierta y nos sirve también para explicar el origen de los cloroplastos, los orgánulos vegetales causantes de la fotosíntesis y del color verde de las hojas. Estos también tienen su propio ADN y fueron, antaño, bacterias de vida libre. Las células vegetales, por tanto, presentan tres ADN: el del núcleo, el de las mitocondrias y el de los cloroplastos.

La maquinaria de traducción de las mitocondrias

Las mitocondrias sorprenden a los científicos desde hace tiempo. Ya en 1979, un estudio genético produjo un resultado inesperado: en las mitocondrias el codón UGA no era leído como codón stop, sino que codificaba para un aminoácido (el triptófano). A posteriori se comprobó también que en las mitocondrias de los vertebrados, los codones AGA y AGG no se leen como arginina, sino como nuevos codones stop; mientras que AUA, que en la célula se lee como isoleucina, dentro de la mitocondria corresponde a una metionina.

La explicación para que dentro de una célula puedan coexistir dos códigos genéticos es que la mitocondria, además de contener su propio ADN, cuenta con su propia maquinaria de traducción, ribosomas particulares y ARNt exclusivos. En ese sentido son organismos independientes, con su código genético particular. Tanto el ADN como los ribosomas de las mitocondrias guardan muchas similitudes con el ADN y los ribosomas de bacterias actuales, lo que refuerza su origen bacteriano.

A lo largo del tiempo se han detectado otras excepciones en el código genético de algunos organismos unicelulares, como leva-

duras y ciertos ciliados, entre ellos el microorganismo marino del género *Mesodinium*. En este, los tripletes UAA y UAG, que en el código estándar funcionan como codones stop, codifican para el aminoácido tirosina. Es importante resaltar que los cambios detectados involucran siempre a codones que tienen sinónimos (como los codones stop), lo que permite que, a pesar de que algunos tripletes cambien de significado, se continúen produciendo proteínas a partir de 20 aminoácidos. Son excepciones que matizan la universalidad del código pero no la contradicen, pues sus principales características, como la lectura por tripletes no solapante y la mayoría de relaciones entre codones y aminoácidos, se mantienen inalteradas en todos los organismos.

Los cambios se restringen a variaciones en el significado de unos pocos tripletes, lo que implica modificaciones de los ARNt de estos orgánulos u organismos. Normalmente, los efectos de este cambio serían devastadores, ya que, de golpe, se cambiaría un aminoácido por otro en muchas proteínas, lo que comprometería su forma y su función. ¿Cómo pueden haberse dado estos cambios en la lectura del ARN y, por ende, en el código genético de unos pocos organismos sin esos efectos nocivos? La respuesta a esta pregunta se halla en la frecuencia de uso de los codones sinónimos en los distintos organismos.

Como sabemos, existen muchos codones ligeramente diferentes que codifican para un mismo aminoácido. Por ejemplo, GGU, GGC, GGA y GGG son sinónimos para glicina y es indiferente tener GGU o GGG en una secuencia de ARN; en ambos casos, la proteína resultante contendrá la misma glicina en la posición adecuada. Si los codones son equivalentes, su frecuencia debería ser la misma, es decir, que en esa secuencia deberían aparecer en un 25% de los casos. Pero no es así: se ha constatado que en cada organismo hay un uso preferencial de los codones sinónimos.

Así, en *Escherichia coli*, una bacteria intestinal ampliamente estudiada, los codones GGC y GGU se usan de manera preferente

para codificar glicina; cada uno de ellos se utiliza en un 35 % de las ocasiones, es decir, que ambos codifican para un 70 % de las glicinas de estas proteínas. Los codones GGA y GGG, por el contrario, solo se utilizan en un 13 % y un 15 % de las veces, respectivamente. Hay diferencias aún más notables, como es el caso de los codones que codifican para la isoleucina (AUU, AUC y AUA). Cada uno de los tres codones debería ser el responsable de un 33 % de las isoleucinas que se añaden a las proteínas y, sin embargo, AUU contabiliza la mitad de ellas, mientras que AUA codifica únicamente para un 10 % de estos aminoácidos.

El uso preferencial de los codones es diferente en cada especie. En humanos, el codón GGU es el menos usado para codificar glicina, a pesar de que en *Escherichia* es uno de los tripletes más utilizados. En la planta del tabaco, en cambio, AUA y AUC se usan en proporción de un 25 % cada uno, a pesar de que AUA era el codón menos utilizado en *Escherichia*.

Las excepciones aparecidas en el código genético podrían derivarse de este uso preferencial de codones. Un cambio en *Escherichia coli* causado por la variación de significado del triplete GGC produciría una catástrofe, pues se modificarían un 35 % de las glicinas del organismo, y este, probablemente, dejaría de ser viable y moriría. En cambio, si el que varía es un codón poco utilizado, como el AUA, solo se verían afectadas un 10 % de las isoleucinas, lo que tendría un impacto menor. Puede que todas estas excepciones sean el resultado de cambios de significado de codones poco utilizados pues, aunque generan alteraciones en las proteínas de los organismos, no comprometen su viabilidad en menor medida.

Algunos resultados recientes parecen indicar que los codones sinónimos podrían no ser estrictamente sinónimos. Algunas posturas señalan que dos codones sinónimos podrían haber dado lugar a dos proteínas con plegamientos diferentes, idénticas entre ellas en un 98 %, pero con funciones ligeramente divergentes. Como sucede con el uso de palabras sinónimas, una diferencia de matiz puede

afectar la expresión del lenguaje, en este caso genético. En otros estudios se apuntaba que los codones sinónimos podrían incidir, entre otras cosas, en la velocidad de transcripción o traducción. Quizá, después de todo, haya una razón evolutiva detrás de esa redundancia de la que hace gala el código genético.

El conocimiento cada vez más profundo que los científicos van adquiriendo del código genético abre la puerta a poder modificarlo a voluntad como, de hecho, se ha conseguido ya en algunos organismos para sintetizar, por ejemplo, proteínas con aminoácidos modificados. Pero como veremos en el próximo capítulo, estos hallazgos no habrían sido posibles sin la labor llevada a cabo por un grupo de brillantes científicos y que permitió el desciframiento del código de la vida.

02

EL DESCIFRAMIENTO DEL CÓDIGO GENÉTICO

El desciframiento del código genético es un logro del criptoanálisis que hubiera enorgullecido al mismísimo Alan Turing. Su descodificación se llevó a cabo en poco más de una década gracias al esfuerzo y la colaboración de científicos de todo el mundo y sentó las bases para la futura lectura y reescritura de la información genética.

A principios de la década de 1940 ya se sabía que los genes pasaban de generación en generación e influían sobre la forma y el comportamiento de los seres vivos. Pero lo que todavía resultaba un misterio era cómo contenían y transmitían la información. Desvelar su naturaleza química y su composición arrojaría luz sobre los numerosos interrogantes que todavía permanecían abiertos e impulsaría definitivamente el estudio de la genética.

En este sentido fue fundamental la aportación del genetista estadounidense Thomas Hunt Morgan, quien, gracias a sus experimentos con moscas de la especie *Drosophila*, consiguió demostrar la vinculación de unas estructuras del núcleo de las células, los cromosomas, con la herencia. Estos están compuestos principalmente por dos tipos de moléculas: proteínas y ácidos nucleicos. Aunque la mayoría de científicos se inclinaba por considerar a las proteínas como las moléculas que transportaban la información celular por su mayor complejidad —estaban formadas por más moléculas diferentes (letras) que el ADN—, diferentes experimentos, como el que en 1944 llevaron a cabo Oswald Theodore

Avery, Colin MacLeod y Maclyn McCarty, acabaron demostrando que era el ADN el que contenía la información genética.

A partir de ese momento, se multiplicaron los trabajos científicos dirigidos a desentrañar su forma y características, trabajos que culminaron con la determinación de la estructura en doble hélice del ADN en 1953. El hallazgo, meramente teórico, aunque basado en datos experimentales obtenidos por otros investigadores, fue publicado en la revista *Nature* el 25 de abril de 1953 por el estadounidense James Watson y el británico Francis Crick. Gracias a él ambos recibieron el premio Nobel en 1962. La estructura de la doble hélice se mantenía por la complementariedad entre las bases del ADN; las adeninas de una cadena siempre se unen a las timinas de la otra cadena; las citosinas de una cadena siempre reconocen a las guaninas. Estas parejas, A:T y C:G, además, presentan un tamaño prácticamente idéntico, lo que hace que la doble hélice se mantenga inalterada sea cual sea su secuencia de nucleótidos; si el tamaño de cada pareja fuera ligeramente diferente, la forma de la doble cadena del ADN dependería de qué parejas la formasen en cada organismo. Pero las dobles cadenas de todos los organismos, con secuencias evidentemente distintas, son idénticas en cuanto a su forma.

EL PAPEL DEL ARN

Este importantísimo descubrimiento evidenciaba la estructura del ADN, pero dejaba sin resolver en qué lenguaje se escribía y leía. Los propios autores del modelo, Watson y Crick, postularon entonces que «parece probable que la secuencia precisa de bases [las letras del ADN] es el código que contiene la información genética». Y añadieron que, quizás, la información del ADN dependía del orden de sus letras, lo que abría la puerta al futuro desciframiento de la pauta de lectura de la base de toda la vida de nuestro planeta que, como

veremos a continuación, dependió en gran medida de las investigaciones llevadas a cabo sobre el ácido ribonucleico (ARN).

La investigación sobre el código genético no hubiera sido posible sin la colaboración de científicos de todo el mundo. Un ejemplo de este ambiente cooperativo fue la creación del Club de las Corbatas de ARN, fundado por el físico y divulgador ucraniano George Gamow, y llamado así porque sus veinte miembros llevaban frecuentemente una corbata con una hélice bordada, que diseñó el propio Gamow. Entre los integrantes de este selecto club destacan los físicos Richard Feynmann y Max Delbrück, el bioquímico Erwin Chargaff o a los propios James Watson y Francis Crick. Seis de los miembros del club acabaron siendo reconocidos con el premio Nobel, aunque ninguno de ellos específicamente por el desciframiento del código.

Así, basándose en el intercambio de información tanto en las reuniones presenciales como mediante la correspondencia postal, el club tenía el objetivo de descifrar un código genético que, como cada vez resultaba más claro, relacionaba el ARN con las proteínas. Eso era así porque durante las dos décadas anteriores a la fundación del club se habían ido acumulando evidencias experimentales sobre el papel que desempeñaba el ARN en la célula. Se había determinado, por ejemplo, que mientras que el ADN se encontraba en el núcleo, el ARN se situaba principalmente fuera de este, en el citoplasma, en el mismo lugar en el que se produce la síntesis de las proteínas. No obstante, algunos artículos también indicaban que el ARN se podía estar sintetizando en el núcleo, lo que lo relacionaría directamente con el ADN. El ARN, por tanto, parecía estar vinculado por un lado con las proteínas y por el otro con el ADN. Más adelante se comprobó que las células que fabrican grandes cantidades de proteínas presentan también cantidades elevadas de ARN y que el uracilo, el nucleótido exclusivo del ARN, era imprescindible para la fabricación de proteínas. A raíz de todos estos resultados, algunos autores postularon que el ARN actuaría como intermediario entre el ADN y las proteínas. De alguna

manera, el ARN se fabricaría en el núcleo a partir del ADN, seguramente gracias a la misma complementariedad entre bases de la doble hélice del propio ADN, y saldría fuera del núcleo donde serviría de molde para la fabricación de proteínas. El código que se quería descifrar era el que relacionaba el idioma de ese ARN con el idioma de las proteínas.

Una de las primeras aproximaciones del Club de las Corbatas al desciframiento del código vino de la mano de Gamow y se basó en cálculos de combinatoria. Como los dos idiomas presentan distinto número de posibles elementos (4 y 20), no puede haber una correspondencia de un elemento con un elemento. Si cada uno de los 4 nucleótidos correspondiese a un aminoácido, quedarían 16 aminoácidos sin correspondencia. Los nucleótidos debían leerse agrupados: quizá de dos en dos, o de tres en tres, o de cuatro en cuatro, etc.

Supongamos que cada aminoácido está codificado por una pareja de nucleótidos, ¿cuántas parejas de nucleótidos podemos formar? Podemos tener AA, AC, AG, AU, CA, CC, CG, CU, GA, GC, GG, GU, UA, UC, UG, UU; es decir, 16 posibles parejas. Un número insuficiente ya que existen 20 aminoácidos; 4 aminoácidos quedarían sin correspondencia. Gamow defendía que las letras del ARN se leían de tres en tres, por tríos o tripletes. ¿Cuántos tríos podemos formar? AAA, AAC, AAG, AAU, ACA, etc. Podemos calcular el número total si consideramos que para la primera posición tenemos 4 posibles letras que se combinan con otras 4 posibles letras en la segunda posición ($4 \times 4 = 16$ parejas) y con otras 4 posibles letras en la tercera posición ($4 \times 4 \times 4 = 64$ tripletes). Los 64 tripletes permiten codificar los 20 aminoácidos (no hay ningún aminoácido sin asignar). Una pauta de lectura por tripletes implicaba que existiesen tripletes sinónimos que codificasen para un mismo aminoácido.

Gamow especulaba, además, que la lectura del ARN era solapante, es decir, que cada «palabra» se leía avanzando una o dos letras. En el primer caso, una misma letra pertenecería a tres tripletes distintos y, en el segundo, una letra pertenecería a dos tripletes

En la imagen superior, George Gamow (derecha) con un colega físico, John Cockcroft. En la inferior, cuatro miembros del Club de las Corbatas del ARN: de izquierda a derecha, Francis Crick, Alexander Rich, Leslie E. Orgel y James Watson.

distintos. La lectura solapante de los tripletes avanzando una sola letra de «Aúnvoysinunauñadelpie» por tripletes daría como resultado la frase «Aún únv nvo voy oys...»; su lectura avanzando dos letras tras leer una palabra nos devolvería la frase «Aún nvo oys sin nun nau...»; la lectura no solapante de la misma secuencia nos devolvería la frase «Aún voy sin una uña del pie».

La lectura solapante implicaba algunas características de las proteínas que no parecían observarse en los experimentos. Así, si suponemos que los tripletes del ARN se solapan en dos letras, una secuencia como AUCGAUCGA se empezaría de tal manera que el primer triplete sería AUC y el segundo contendría las dos últimas letras (UC) del primer triplete más una letra nueva (G), dando lugar a un segundo triplete (UCG). Como esta nueva letra solo podría ser A, C, G o U —ya que no hay más letras en el ARN—, al aminoácido determinado por AUC, fuese el que fuera, siempre le seguiría un aminoácido de entre cuatro posibles (codificados por UCA, UCC, UCG o UCU). Los aminoácidos de la secuencia de una proteína estarían, en parte, determinados por el aminoácido anterior. Por aquella época se había empezado a leer la secuencia de aminoácidos de diferentes proteínas y se comprobó que no había agrupaciones de aminoácidos preferentes; después de un aminoácido se podía encontrar cualquiera de los 20 aminoácidos y no solo 4 posibilidades. Un aminoácido no condicionaba a los aminoácidos que le seguían. Además, según el modelo de la lectura solapante, la mutación de una sola letra del ARN afectaría a tres aminoácidos. Así, en la secuencia AUCGAUCGA la mutación de la primera C por una U (AUUGAUCGA) afectaría a tres tripletes (AUU, UUG y UGA) y, por tanto, a tres aminoácidos. Las evidencias que se fueron acumulando parecían indicar que una mutación en una letra del ARN afectaba a un solo aminoácido. A raíz de los experimentos, la lectura solapante fue perdiendo fuerza y se descartó definitivamente en 1961 (fig. 1).

Desde la década de 1950 se conocía por los trabajos del biólogo estadounidense Paul Zamecnik que las proteínas se sintetizan en

Fig. 1

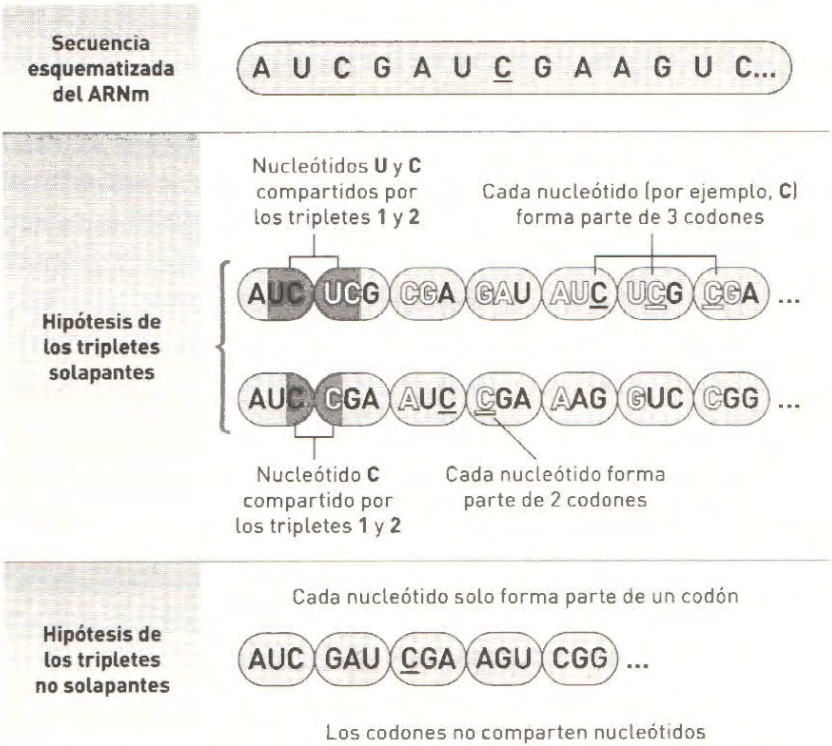


Figura en la que se muestran los patrones de lectura sobre una secuencia de ARN. En la parte central, las lecturas por tripletes solapantes en dos o en un nucleótido; abajo, la lectura por tripletes no solapantes.

partículas riboproteicas, que posteriormente se bautizaron con el nombre de *ribosomas*. Por lo tanto, estos tenían que participar en alguna etapa de la traducción de la información genética a proteínas.

Por otro lado, diversos experimentos demostraron que, al añadir una enzima que degradaba el ARN, se detenía la síntesis de proteínas. El ARN se iba consolidando como la molécula que participa-

ba directamente en la síntesis de las proteínas. ¿Cómo encajar esta nueva pieza en el puzle genético?

La respuesta vino de nuevo de la mano de Francis Crick, quien, en 1958, propuso una serie de hipótesis que iban a cambiar radicalmente la manera de abordar el estudio del código genético. La primera, ya indicada anteriormente, hacía referencia al modo en que circula la información por las tres grandes moléculas (ADN, ARN y proteínas). Bautizada como el *dogma central de la biología*, esta hipótesis postula que el flujo de información en los sistemas biológicos va siempre en la misma dirección: el ADN se replica antes de que la célula se divida, con el fin de dotar a cada célula hija de una copia idéntica del material genético; después, una de sus cadenas —llamada *cadena codificante*— se transcribe a ARN, y finalmente este se traduce a proteínas. Aunque con algunas modificaciones posteriores, la idea básica de este dogma sigue siendo vigente hoy en día.

> Una buena teoría no solo predice, sino que realiza predicciones sorprendentes que acaban resultando ciertas.
>
> FRANCIS CRICK

Crick también dedujo, y de nuevo con acierto, que la secuencia del ADN determinaba el orden de las letras de la proteína y que esta secuencia lineal de la proteína determinaba su plegamiento y, por tanto, su función. Además, propuso la existencia de una molécula que actuaría de intermediaria entre el ARN y las proteínas.

Hasta entonces, una de las teorías que se barajaban era la de que el ARN actuaba como molde físico y real de las proteínas. Una combinación de nucleótidos del ARN dejaría un hueco en el que encajaría un aminoácido concreto. Crick, junto con otro miembro del club, el biólogo sudafricano Sydney Brenner, había estado analizando la forma conocida de los aminoácidos y no encontraba manera de que encajasen físicamente con los nucleótidos. Este encaje inexistente no podía, por tanto, determinar su relación. Debía exis-

tir una molécula que funcionara de adaptador entre la molécula de ARN y el aminoácido que se añadiría a la proteína que se sintetiza. En palabras del propio Crick:

> La idea fundamental era que resulta muy difícil considerar cómo el ADN o el ARN en cualquier posible estructura pueden constituir un molde inmediato para las cadenas laterales de los veinte aminoácidos. Lo que se esperaría de cualquier estructura posible es una distribución específica de grupos atómicos que puedan formar enlaces de hidrógeno. Por lo tanto, propuse una teoría según la cual existirían veinte adaptadores (uno para cada aminoácido), junto con veinte enzimas especiales. Cada enzima sería capaz de unir un aminoácido concreto a su propio adaptador específico. Estos complejos se trasladarían por difusión al molde de ARN. Una molécula de adaptador podría encajar solo en aquellos lugares del molde de ácido nucleico en los que pudiera formar los enlaces de hidrógeno necesarios para mantenerlo en su lugar. Al hacer eso, el adaptador habría llevado a su aminoácido precisamente al lugar correcto en que hacía falta.

Apenas medio año después de esta formulación, el grupo de Zamecnik describió la existencia de unas moléculas de ARN que desempeñaban el papel sugerido por Crick. Puesto que su función es transferir el aminoácido que transporta a la proteína que se está sintetizando, esa nueva especie de ARN se denominó ARNt.

EL MENSAJERO

Si el ARNt era el adaptador, ¿qué ARN contiene la información genética? En un principio se pensó que podía ser el ARN que se encuentra en los ribosomas, ya que era el único otro ARN claramente identificado y que, además, se hallaba fuera del núcleo,

asociado a proteínas que se estaban sintetizando. Según esta hipótesis, cada ribosoma sería diferente ya que contendría, en su ARN, la información para sintetizar una proteína concreta. Sin embargo, no todos los datos cuadraban en esta hipótesis.

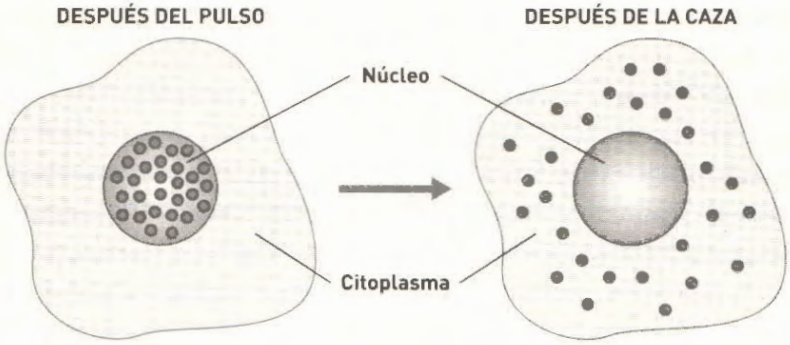
En 1956 Elliot Volkin y Lawrence Astrachan habían publicado un artículo en el que se hablaba de una pequeña fracción de ARN que se sintetizaba tras la infección de bacterias con un fago y que, además, resultaba tener una composición de bases similar a la del fago y no a la de la bacteria huésped. Podría tratarse de un intermediario entre ADN y proteínas, pero Volkin y Astrachan, conscientes de las posibles limitaciones de su experimentos, no descartaron que los resultados pudieran provenir de algún tipo de contaminación.

La similitud de resultados tras la repetición de los experimentos y la comprobación de que la secuencia del ARN transitorio era complementaria a la secuencia de un fragmento del ADN y el hecho de que ambas moléculas podían hibridar, unirse, para formar una molécula de ADN-ARN hicieron que su importancia dentro de los procesos biológicos empezase a considerarse como una posibilidad real.

El impulso definitivo se produjo en 1960, durante una reunión informal entre Sydney Brenner, Francis Crick y François Jacob, miembro de un importante grupo de investigadores de París que estudiaban la regulación de los genes en bacterias. Jacob compartió con sus colegas el resultado de uno de los últimos experimentos que se habían llevado a cabo en su laboratorio. En él se había marcado un gen bacteriano con fósforo radiactivo que luego, a través de la desintegración del fósforo, se destruía. El resultado era claro: tras destruir el gen no se encontraba enzima. Cuando faltaba el gen, la síntesis de proteínas se detenía, lo que excluía que el intermediario entre el ADN y las proteínas fuera un intermediario estable, pues este hubiera podido mantenerse en ausencia del gen.

> EXPERIMENTOS DE PULSO Y CAZA

Los estudios que condujeron al desciframiento del código y al papel que desarrollaba el ARN en el flujo de información genética no se podrían haber realizado en aquel momento sin el uso de materiales radiactivos, los cuales permiten seguir las moléculas construidas a partir de precursores marcados radiactivamente. En los denominados *experimentos de pulso y caza* se incuban las células en un medio con un componente marcado radiactivamente como, por ejemplo, el nucleótido uracilo. Esta primera incubación se conoce como *pulso*. Después, se incuban las mismas células en un medio sin ningún componente radiactivo para observar dónde se ha utilizado la molécula radiactiva del medio anterior; es decir, se «caza» la molécula radiactiva. En el caso de las células incubadas con uracilo radiactivo («pulso»), cuando se pasan a un medio sin radiactividad y se sigue el rastro de la radiactividad («caza»), se observa como justo después del pulso, el uracilo se sitúa en moléculas de ARN dentro del núcleo (se está transcribiendo ARN con el uracilo radiactivo). Pasado un tiempo corto, el ARN radiactivo sale al citoplasma, donde se degrada. Estos experimentos permitieron establecer dónde se sintetiza el ARN y constatar que había tipos de ARN de vida muy corta que se degradan rápidamente.



— En el experimento de pulso y caza con uracilo radiactivo, el ARN radiactivo deja el núcleo de la célula para salir al citoplasma, donde se degrada.

Ante estos resultados, Crick y Brenner apuntaron la posibilidad de que este ARN fuera el intermediario real entre el ADN y las proteínas. Para confirmarlo, Brenner y Jacob decidieron que cultivarían bacterias en un medio con isótopos pesados, las infectarían con fagos y después las pasarían a un medio con isótopos ligeros al que añadirían fósforo radiactivo. Tras eso, separarían los ribosomas bacterianos en un gradiente de cloruro de cesio y, si sus deducciones eran ciertas, el pico de radiactividad debería asociarse únicamente con los ribosomas pesados, los preexistentes.

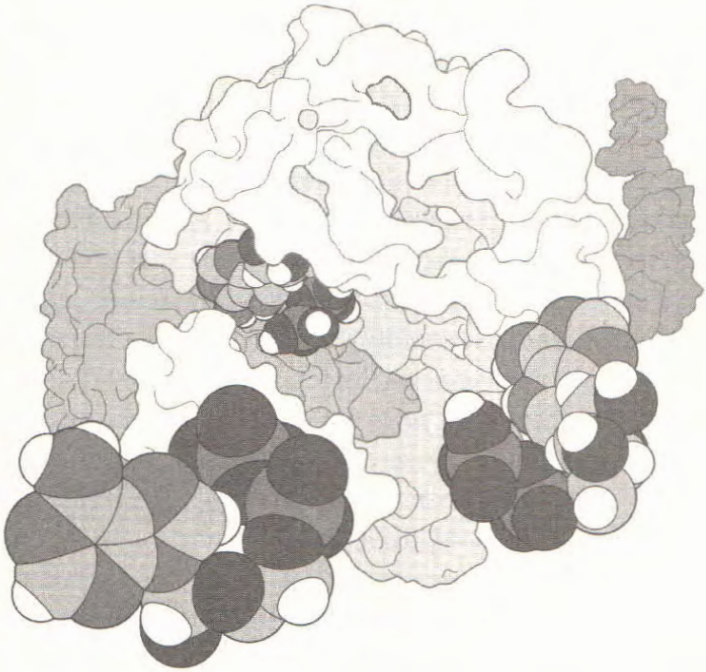
Tras unos meses en el laboratorio de Matthew Meselson pusieron a punto las condiciones y confirmaron su teoría: habían identificado el ARNm. Los ribosomas no llevaban, pues, ninguna información sobre las proteínas que sintetizaban, sino que se limitaban a leer el ARNm. Si se quería descifrar el código genético, se tenía que descifrar, concretamente, el código del ARNm.

La clave vino de la mano de un joven investigador estadounidense, Marshall Nirenberg, quien, en 1961 y en colaboración con el alemán Heinrich Matthaei, expuso unos resultados que dejaron asombrado a Francis Crick hasta el punto de afirmar que con dicho descubrimiento «estamos llegando al final de una era en la biología molecular. Si la estructura del ADN fue el fin del principio, el descubrimiento de Nirenberg y Matthaei es el principio del fin».

LA PRIMERA LETRA

Los experimentos que Nirenberg presentó en 1961 se basaban en el uso de una enzima, la polinucleótido fosforilasa (fig. 2), descubierta por el científico español Severo Ochoa y que le valió el premio Nobel en 1959. Esta enzima, dentro de la célula, desmonta el ARN pieza a pieza, es decir, va desgranando una hebra de ARN liberando sus nucleótidos. La mayoría de las reacciones que tienen lugar en nuestras células, por no decir todas, son reversibles, es decir, pue-

Fig. 2



Estructura de la proteína polinucleótido fosforilasa incorporando nucleótidos para sintetizar ARN. Esta enzima, descubierta por Severo Ochoa, permitió los primeros experimentos que descifraron el código genético.

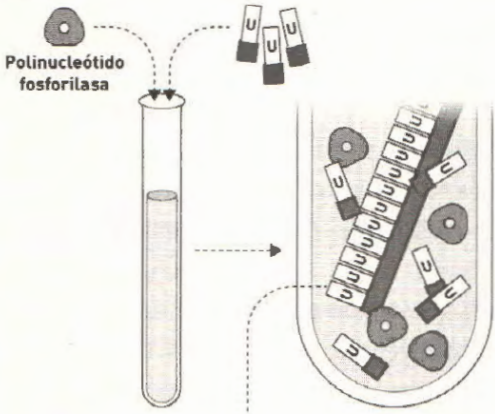
den ocurrir en un sentido o en el contrario. Si se modifican algunas variables, se puede conseguir que una enzima participe en la reacción inversa. Si una reacción une A y B para formar C, se pueden modificar las condiciones para, a partir de C, obtener A y B. En el caso de la polinucleótido fosforilasa identificada por Ochoa, en condiciones controladas, se conseguía que la enzima sintetizase una cadena de ARN a partir de nucleótidos sueltos. Como la enzima fabrica el ARN sin ningún molde, incorpora las piezas (los nucleótidos de A, C, G y U) al azar.

La idea que perseguían tanto Nirenberg como Ochoa —y que en esa carrera ganaría el primero— consistía en utilizar dicha enzima pero ofreciéndole solo un tipo de nucleótidos, por ejemplo, los de uracilo. En este caso, la enzima sintetizaría largos ARN formados por repeticiones de nucleótidos de uracilo: UUUUUUUUUUUU... Nirenberg y Matthaei habían estudiado y perfeccionado un sistema de producción de proteínas sin células que consistía en una mezcla de los elementos necesarios para sintetizar una proteína y contenía, entre otros, ribosomas, ARN y aminoácidos. Algunos de los aminoácidos estaban marcados radiactivamente para poder seguir la producción de proteínas. En sus experimentos habían comprobado que la síntesis de proteínas se iniciaba cuando añadían ARN «plantilla» o «modelo», es decir, el ARNm, aunque el uso de esta denominación aún no se había extendido. La síntesis proteica se producía tanto si el ARNm procedía de una célula como si se había sintetizado en un tubo de ensayo a partir de la enzima purificada por Ochoa. Cuando los investigadores añadieron la cadena UUUUUUUU... al sistema de producción, obtuvieron una proteína enteramente formada por fenilalanina.

Los investigadores realizaron una gran cantidad de experimentos para corroborar sus resultados. En cada experimento modificaban algunas de las variables y medían la radiactividad del resultado. La proteína formada a partir de UUUUUUUU... solo era radiactiva si el aminoácido radiactivo que se añadía al medio era la fenilalanina; si se añadían otros aminoácidos radiactivos, pero la fenilalanina añadida no era radiactiva, la proteína no era radiactiva. Si en vez de UUUUUUU... se añadían cadenas de ARN con otros nucleótidos repetidos (AAAAA..., CCCCC... o GGGGG...) y se cultivaban con fenilalanina radiactiva, la proteína resultante no era radiactiva, puesto que incorporaba otros aminoácidos distintos a la fenilalanina (fig. 3).

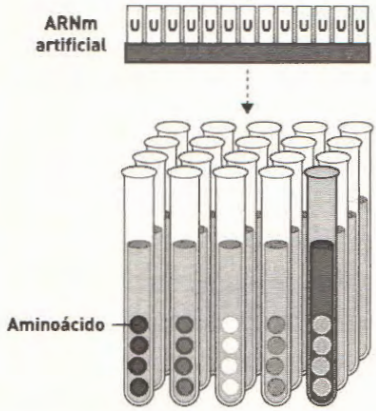
Este fue el resultado que sorprendió tanto a Crick: se había establecido la primera relación unívoca entre una secuencia de ARN,

Fig. 3



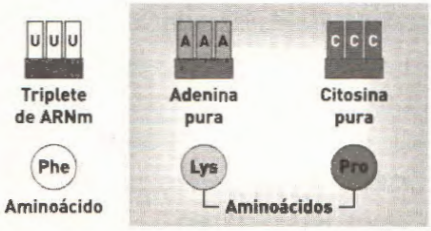
Síntesis del ARN

Incubando la polinucleótido fosforilasa en presencia, únicamente, de uracilo (U), se obtienen largas cadenas de ARN formadas exclusivamente con este nucleótido, denominadas *cadenas poli(U)*.



Síntesis de proteínas

Se añade el poli(U) a diferentes tubos de ensayo, cada uno de ellos con un aminoácido concreto radiactivo y el resto no radiactivos para comprobar qué aminoácidos incorporará la proteína sintetizada a partir del poli(U).



Lectura de la radiactividad

Solo era radiactiva la proteína del tubo en el que la fenilalanina era radiactiva. Experimentos posteriores similares permitieron establecer que AAA codifica para lisina y CCC para prolina.

Representación del experimento de Nirenberg y Matthaei que permitió establecer las primeras relaciones del código genético.

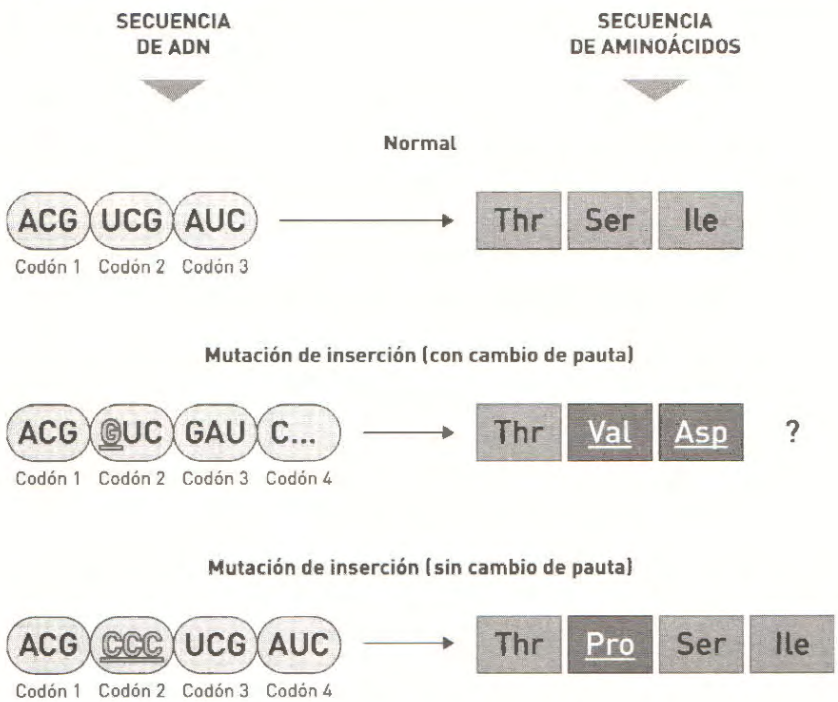
la repetición de nucleótidos de uracilo, y un único aminoácido, la fenilalanina. El código empezaba a descifrarse, y lo habían conseguido dos personas ajenas al Club de las Corbatas del ARN.

El experimento de Nirenberg, sin embargo, no establecía cómo se pasaba de una repetición de nucleótidos de uracilo a una repetición de aminoácidos fenilalanina; es decir, no demostraba cómo se leía el código. Un artículo publicado en 1961 por Francis Crick contenía las respuestas.

Para responder a la pregunta de cómo se leen, por ejemplo las U, ¿de una en una, de dos en dos, de tres en tres?, Crick y sus colaboradores provocaron una mutación en la información genética de un virus con unas sustancias que añaden o quitan nucleótidos a las secuencias de su genoma. Cuando añadían o quitaban un nucleótido o una pareja de nucleótidos, la secuencia de aminoácidos de la proteína sintetizada cambiaba completamente a partir del lugar de la mutación. Por ejemplo, si a la secuencia ACGUCGAUC..., que se traduce como treonina-serina-isoleucina-..., se le añadía una G o una pareja AC en cuarta posición obtenían las secuencias ACGGUCGAUC... y ACGACUCGAUC..., que se traducen, respectivamente, a treonina-valina-ácido aspártico-..., y a treonina-treonina-arginina-... En ambos casos, los aminoácidos a partir de la primera treonina y hasta el final de la proteína están completamente cambiados respecto de la secuencia original. Sin embargo, cuando se añadía o quitaba un grupo de tres nucleótidos, aparecía o desaparecía un aminoácido, pero el resto de la secuencia continuaba inalterada. Si a la secuencia ACGUCGAUC... se le añade el trío de letras CCC en cuarta posición, se obtiene la secuencia ACGCCCUCGAUC... que se traduce como treonina-prolina-serina-isoleucina-... La nueva proteína contiene una prolina adicional en segunda posición, pero el resto de aminoácidos son idénticos a los de la secuencia original.

Las inserciones o deleciones de 1 o 2 nucleótidos cambiaban drásticamente la pauta de lectura del ARN; si se añadían o quitaban 3 nucleótidos, la pauta de lectura se mantenía (fig. 4).

Fig. 4



Efectos de las inserciones que afectan (centro) o que no afectan (inferior) a la pauta de lectura.

Ello llevó a la conclusión de que el ARNm se lee por tripletes, pues si los nucleótidos de ARN se leyeran en grupos de dos o de cuatro aminoácidos, los resultados para cada mutación habrían sido distintos.

Crick también estableció que el código se lee de forma no solapante, con lo que el cambio de una letra afecta a un solo triplete y, por tanto, a un solo aminoácido. Cada ARN, además, se empieza a leer desde el mismo lugar de inicio y no existen letras que separen los tripletes, no hay comas entre las palabras. Finalmen-

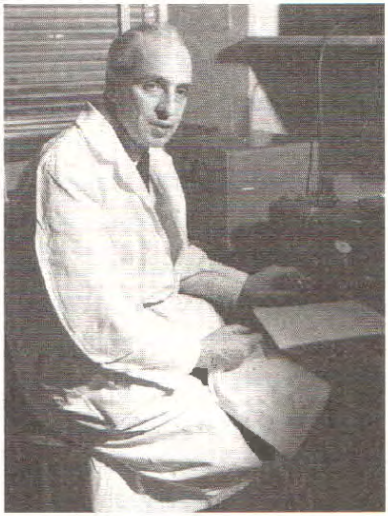
te, los autores proponían que el código es redundante, puesto que al leerse por tripletes hay más tríos de nucleótidos, 64, que aminoácidos, 20, lo que implica que diferentes tripletes codifican para un mismo aminoácido. El artículo confirmaba dos postulados de Gamow, la lectura por tripletes y la redundancia del código, pero refutaba otro, la lectura solapante.

En las conclusiones del artículo Crick añadió un reconocimiento al trabajo de Nirenberg y propuso que, según sus resultados, la fenilalanina estaría probablemente codificada por un triplete de nucleótidos de uracilo. La primera correspondencia del código genético había quedado establecida. El artículo se cerraba con la siguiente reflexión:

> Es posible, por diferentes métodos químicos o enzimáticos, sintetizar polirribonucleótidos [cadenas de ARN] con una secuencia definida o parcialmente definida. Si estos, a su vez, producen polipéptidos específicos, el problema del código está abierto de par en par para ser atacado experimentalmente, y de hecho muchos laboratorios, incluyendo el mío, ya están trabajando en ello. Si la lectura del código es la que nuestros resultados sugieren, y el código es universal, entonces el código genético podría estar resuelto en un año.

Esta previsión, aunque resultó ser optimista, dejaba claro que el desciframiento era solo una cuestión de tiempo; un tiempo, además, muy corto.

Mientras se iniciaba el desciframiento del código del ARN, otros investigadores, como el bioquímico estadounidense Robert Holley, habían centrado su atención sobre el ARNt. Sus investigaciones contribuyeron a demostrar que se trataba de moléculas de ARN unidas cada una de ellas a un aminoácido específico. La determinación de su estructura y su secuencia permitieron establecer que todos los ARNt presentaban la misma forma, con un aminoácido en un extremo de la molécula y tres nucleótidos (anti-



— Arriba, tres de los implicados en la carrera por descifrar las combinaciones posibles del código genético: a la izquierda, Heinrich Matthaei, de pie, con Marshall Nirenberg; a la derecha, Severo Ochoa. En la imagen inferior, los tres agraciados con el premio Nobel de 1968, con Nirenberg entre Robert Holley (izquierda) y Har Gobind Khorana «por su interpretación del código genético y su función en la síntesis de proteínas».

codón) en un bucle de la misma que son complementarios al triplete al que se unen. El ARNt que lee UUU (codón), por ejemplo, presenta una secuencia AAA (anticodón) en un determinado bucle, que es complementaria al UUU que reconoce, y una fenilalanina en el extremo contrario.

EL DESCIFRADO PROBABILÍSTICO

Tras el éxito con el triplete de uracilo, se sintetizaron cadenas de ARN que solo contenían o bien A, o bien C o bien G. El ARN solo con A (AAAAAA...) sintetizaba una proteína formada exclusivamente por lisinas; las cadenas CCCCCC... producían proteínas con solo prolinas; los resultados con cadenas GGGGGG..., aunque esquivos al principio, mostraron que GGG codificaba para la glicina. Se habían establecido, por tanto, tres nuevas relaciones del código genético: AAA y lisina; CCC y prolina; GGG y glicina.

Las repeticiones de un solo nucleótido son fáciles de obtener, únicamente se necesita cultivar la enzima polinucleótido fosforilasa descubierta por Severo Ochoa con una mezcla de nucleótidos en la que haya nucleótidos de un tipo. ¿Qué pasa, sin embargo, cuando se cultiva la enzima con una mezcla de nucleótidos de dos tipos distintos? Recordemos que, al no necesitar molde, la enzima fabrica el ARN uniendo las piezas al azar. Si la cultivamos con nucleótidos de uracilo y citosina, por ejemplo, la enzima genera una cadena en la que alterna de manera aleatoria las dos letras. Si se sintetizan proteínas a partir de los ARN resultantes de esta mezcla de U y C, se obtienen proteínas con otros aminoácidos como prolinas, glutaminas, histidinas, treoninas, asparaginas y lisinas. En aquella época no había ninguna manera de leer la cadena de nucleótidos que se generaba, con lo que establecer las relaciones entre los tripletes formados por dos letras distintas y los aminoácidos correspondientes requería un refinamiento de la aproximación que incorporase cálculos probabilísticos.

Si tenemos una mezcla con la misma cantidad de U y C y generamos tripletes al azar, hay un 50 % de probabilidades de incorporar un U y un 50 % de incorporar una C en la primera posición. Dicho de otra manera, de cada 100 veces que escojamos una letra es de esperar que 50 veces corresponda a una U. En probabilidad, en vez de referirnos a un conjunto de 100 casos, los cálculos se realizan asumiendo un único caso; podemos obtener este valor dividiendo entre cien los valores con el tanto por ciento. Así, para la primera posición tenemos una probabilidad de 0,5 de encontrar una U y una probabilidad de 0,5 de encontrar una C. La probabilidad para las siguientes posiciones es exactamente la misma. Podemos calcular las probabilidades de obtener cada triplete multiplicando la probabilidad de encontrar cada nucleótido en una posición concreta. Así, la probabilidad que tiene la enzima de fabricar un triplete UCU en estas condiciones es de 0,125, que resulta de multiplicar 0,5 para la primera U por 0,5 para la C en segunda posición y por 0,5 para la última U. Si utilizamos una mezcla de U y C en proporciones iguales, tenemos una probabilidad del 12,5 % de obtener el triplete UCU. En la tabla siguiente se pueden observar las probabilidades para todos los tripletes resultantes de esta mezcla que, en este caso, son iguales para todos ellos.

Triplete	Cálculo	Probabilidad	Porcentaje
UUU	0,5×0,5×0,5	0,125	12,5
UUC	0,5×0,5×0,5	0,125	12,5
UCU	0,5×0,5×0,5	0,125	12,5
UCC	0,5×0,5×0,5	0,125	12,5
CUU	0,5×0,5×0,5	0,125	12,5
CUC	0,5×0,5×0,5	0,125	12,5
CCU	0,5×0,5×0,5	0,125	12,5
CCC	0,5×0,5×0,5	0,125	12,5
	Total	**1**	**100**

¿Qué ocurre, sin embargo, si en la mezcla de nucleótidos hay más de un tipo que de otro? Supongamos que preparamos una mezcla en la que hay un 60 % de C y un 40 % de U. Las probabilidades de coger, al azar, una C en cualquiera de las posiciones son ahora del 60 % (p = 0,6); y del 40 % (p = 0,4) para la U. Así, la probabilidad de obtener UCU es de 0,096, que resulta de multiplicar una probabilidad de 0,4 de obtener la primera U, por la probabilidad 0,6 de obtener la C, por el 0,4 de la última U. Si utilizamos una mezcla del 60 % de C y el 40 % de U, tenemos una probabilidad del 9,6 % de encontrar una UCU. Las probabilidades para los posibles tripletes quedaría como sigue:

Triplete	Cálculo	Probabilidad	Porcentaje
UUU	0,4×0,4×0,4	0,064	6,4
UUC	0,4×0,4×0,6	0,096	9,6
UCU	0,4×0,6×0,4	0,096	9,6
UCC	0,4×0,6×0,6	0,144	14,4
CUU	0,6×0,4×0,4	0,096	9,6
CUC	0,6×0,4×0,6	0,144	14,4
CCU	0,6×0,6×0,4	0,144	14,4
CCC	0,6×0,6×0,6	0,216	21,6
	Total	**1**	**100**

Modificando la proporción en la mezcla inicial de nucleótidos se obtienen ARN cuyos tripletes tienen una proporción determinada y distinta entre sí, hecho que debería verse reflejado en la proporción de los aminoácidos que formarán la proteína a partir de ese ARN. El equipo de Nirenberg realizó el experimento mezclando U y C en la proporción de 60:40 que nos ha servido para calcular la tabla y obtuvo proteínas que analizó para comprobar de qué aminoácidos estaban formados y cuál era su proporción. En ese experimento en concreto se obtuvieron proteínas con un 45,7 % de prolinas, un 25,6 % de serinas, un 16 % de fenilalaninas y un 12,7 % de

leucinas. *A priori* estos valores no encajan con los valores de los tripletes de la tabla, pero ofrecen algunas pistas. Así, por ejemplo, se sabía que UUU codificaba para fenilalanina, con lo que se esperaba que la fenilalanina apareciera en una proporción igual o mayor a la proporción de este triplete (6,4 % en este caso). La proporción de fenilalanina es del 16 %, mayor que el 6,4 % de UUU. Por lo tanto, debería existir otro triplete que codificara para fenilalanina y, probablemente se trate de un triplete con dos U y una C (UUC, UCU o CUU), ya que cualquiera de ellos presenta una proporción de 9,6 %, la cual sumada a la proporción de 6,4 % de UUU nos da el 16 % de la fenilalanina. Este experimento, sin embargo, no nos indica cuál de los tres tripletes candidatos codificaba efectivamente para fenilalanina. Los investigadores también tenían que tener en cuenta las desviaciones de los resultados debidas al propio azar y al error asociado a cualquier resultado experimental. De este modo, el 12,7 % de la leucina podía corresponder a un triplete que tuviera tanto una como dos C (con un 9,6 % y un 14,4 %, respectivamente).

Los resultados de cada uno de estos experimentos no señalaba de manera inequívoca prácticamente ninguna relación entre tripletes y aminoácidos. Fue el análisis combinado de los resultados de toda una serie de experimentos similares, en los que se modificaron las bases mezcladas y sus proporciones, el que condujo al desciframiento de buena parte del código genético. El desciframiento completo se estaba resistiendo y no se había producido en un año, como habían propuesto Crick y sus colaboradores. Además, en 1962, empezaban a aparecer codones, es decir, tripletes, que no se ajustaban a lo esperado, tripletes «sin sentido».

Se comprobó que algunas mutaciones provocaban que las proteínas quedaran truncadas, ya que provocaban que, a partir de la posición en la que estaban dichas mutaciones, no se añadiesen ni se quitasen nuevos aminoácidos. Eran mutaciones puntuales «sin sentido». ¿Cómo podía un cambio de letra provocar no un cambio de aminoácido, sino un cambio en el tamaño de la proteína sinte-

tizada? Con el tiempo se comprobó que la respuesta se encontraba en que algunos tripletes no codifican para ningún aminoácido, sino que indican al ribosoma que la síntesis de proteína debe finalizar. El primer triplete «sin sentido» que se describió en 1965 recibió la denominación de *ámbar*, por la traducción del apellido Bernstein (piedra caliente) de un amigo alemán de los investigadores que lo describieron (Brenner, Stretton y Kaplan). Esto condicionó el nombre de los siguientes tripletes «sin sentido» que se identificaron, denominándoles *ocre* y *ópalo*. Al poco de ser descritos los tres, se estableció su secuencia: UAG, UAA y UGA. Como son tripletes que al ser leídos provocan la parada de la traducción, se agruparon bajo el apelativo de codones stop. Pocos años más tarde, se demostró que también existía un codón para señalar el inicio de la traducción. El codón inicio AUG, a diferencia de los stop, codificaba también para un aminoácido, la metionina. Se trata de un codón polisémico, ya que tiene dos significados: indicar el inicio de la síntesis y codificar la metionina.

NUEVAS TÉCNICAS

El impulso definitivo al desciframiento completo del código vino de la mano de la aplicación de nuevas técnicas experimentales, como la bautizada como *polinucleótidos de secuencia alternante* que consistía en la construcción de cadenas de ADN a partir de la repetición de bloques de 2, 3 o 4 nucleótidos. Así, a partir de la repetición del bloque TC, se obtendría la cadena TCTCTCTCTCTCT...; a partir del bloque CCA, obtendríamos la secuencia CCACCACCACCACCA...; o a partir del bloque TCTA, obtendríamos el ADN TCTATCTATCTATCTATCTA... Estas cadenas de ADN, por métodos enzimáticos, se transcribirían a moléculas de ARN con secuencias conocidas ya que deben ser complementarias a las secuencias del ADN, lo que supuso un gran avance respecto a las cadenas de ARN de secuencia aleatoria

utilizadas en los experimentos anteriores. De hecho, se consideró un avance tan crucial que el mayor experto en la técnica, Har Gobind Khorana, compartió el premio Nobel de 1968 con Nirenberg y Holley.

> El progreso de la ciencia depende de la aparición de técnicas nuevas, descubrimientos nuevos e ideas nuevas, seguramente en este orden.
>
> SYDNEY BRENNER

Cuando se traducían las cadenas de ARN de secuencia conocida se obtenían proteínas con una composición de aminoácidos específica que en algunos casos corroboró y en otros desentrañó las relaciones entre tripletes y aminoácidos. Así, la secuencia UCUCUCUCUCUC..., leída como UCU CUC UCU CUC..., codificaba para proteínas en las que se alternaba la serina, codificada por UCU, con la leucina, codificada por CUC. La secuencia CCACCACCACCACCA... producía tres tipos de proteínas, cada una de ellas formada exclusivamente por la repetición de un solo aminoácido. Así, se obtenían proteínas formadas solo por prolinas, solo por histidinas o solo por treoninas. Las diferencias se debía a la letra en qué empezara la traducción. Si empezaba en la primera C, el código se leía como CCA CCA CCA... y todos los aminoácidos eran prolinas; si empezaba en la segunda C, el código se leía como CAC CAC CAC... y las proteínas resultantes solo tenían histidinas; si empezaba por la A, los tripletes eran todos ACC ACC ACC... y los aminoácidos, treoninas. La secuencia TCTATCTATCTATCTATCTA..., por su parte, producía proteínas formadas por la repetición de una serie de serina (TCT), isoleucina (ATC), tirosina (TAT) y leucina (CTA).

Otra nueva aproximación que sirvió para descifrar definitivamente el código genético se basaba en los estudios, cada vez más detallados, del mecanismo molecular de la traducción. Se había comprobado que el ARNm es leído, triplete a triplete, por el ribosoma. Al leer un nuevo triplete, el ribosoma se detiene y el ARNt se une al ribosoma y, por complementariedad, al mensajero; una vez

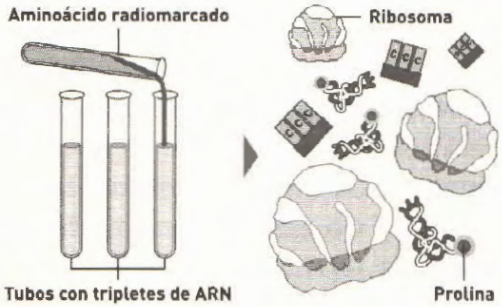
allí, el aminoácido que transporta el ARNt se incorpora a la proteína que se está sintetizando. Este instante de tiempo en el que el ARNt, aún con el aminoácido a cuestas, se une a su triplete en el ribosoma fue utilizado para «observar» directamente las relaciones entre tripletes y aminoácidos.

Un experimento del propio Nirenberg demostró que una cadena muy corta de ARN de solo 3 nucleótidos era suficiente para producir el reclutamiento del ARNt adecuado. Un simple triplete de ARN era capaz de actuar como ARNm y provocar el ensamblaje de toda la maquinaria de traducción para conseguir la «síntesis» de una proteína de un solo aminoácido. Estos tripletes sueltos se podrían utilizar para «pescar» los ARNt que les reconocían y que llevaban asociados el aminoácido correspondiente. Si los tripletes eran el anzuelo, los ribosomas eran la caña. Si se mezclaban ribosomas, tripletes sueltos de ARN y ARNt, estos morderían el anzuelo de los tripletes y quedarían detenidos en los ribosomas. ¿Cómo identificar qué pez se había pescado? Para eso se utilizaba la radiactividad. Si en cada mezcla se utilizaba un solo tipo de aminoácido marcado radiactivamente, se podía comprobar qué aminoácido se unía a cada secuencia. Únicamente faltaba un método para separar los ARNt unidos al ribosoma de los ARNt que no se le habían unido y que continuaban libres en el medio. Para ello, se utilizó un colador. El principio de un colador, de un filtro o de un tamiz es simple: todo lo que sea mayor que el tamaño del poro quedará atrapado. En los experimentos se utilizaron filtros de nitrato de celulosa que pueden tener poros menores de una micra. Los ribosomas, grandes agregados de proteínas y ARN quedan retenidos en el filtro; los ARNt, de tamaño más modesto, pasan por los poros. Se trataba de una pesca con anzuelo (los tripletes de ARNm), caña (el ribosoma) y tamiz (los filtros) en la que el pez que se quiere pescar es el aminoácido que lleva el ARNt (fig. 5).

Supongamos que construimos moléculas de ARN de solo tres nucleótidos con la secuencia GUU. Preparamos distintos reci-

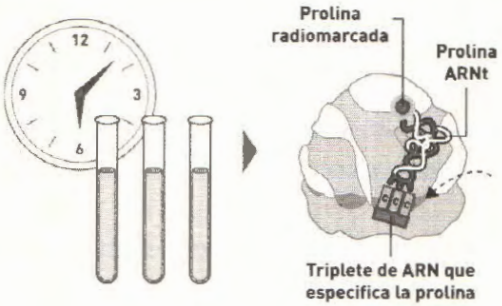
Fig. 5

NIVEL EXPERIMENTAL | NIVEL CONCEPTUAL



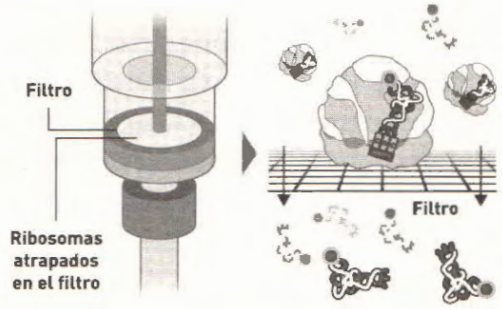
Mezcla inicial

A unos tubos con ribosomas y tripletes de ARNm de secuencia conocida se les añade un ARNt específico para cada tubo, el cual lleva un aminoácido marcado radiactivamente.



Reconocimiento

En los ribosomas el triplete de ARNm es reconocido por el ARNt correspondiente. Si este transporta un aminoácido radiactivo, el complejo que se forma será radiactivo.



Filtrado

Se usa un filtro que atrapa a los ribosomas, pero deja pasar a los ARNt libres. Si se detecta radiactividad, el triplete de ARNm codifica para el aminoácido radiactivo de la mezcla.

Representación del experimento de Nirenberg en el que se comprueba la relación entre tripletes y aminoácidos atrapando con un filtro los ribosomas que unen ambos elementos.

pientes y en todos ellos mezclamos los tripletes GUU con ribosomas. A cada uno de ellos le añadiremos un ARNt con un aminoácido marcado radiactivamente. Para simplificar, imaginemos que probamos dos aminoácidos. En el primer recipiente añadimos un ARNt con prolina marcada radiactivamente, incubamos la mezcla y la filtramos. Si al analizar el filtrado vemos que no detectamos radiactividad eso nos estará indicando que los ARNt con los aminoácidos radiactivos han cruzado el filtro. Los ribosomas que habrían quedado atrapados en el filtro, pese a que deberían haber incorporado el aminoácido correspondiente al ARNt que es complementario con el triplete GUU, no indica señal radiactiva de la prolina, lo que permite concluir que el triplete GUU no codifica para la prolina. Si repetimos el mismo proceso añadiendo esta vez un ARNt con valina marcada radiactivamente, el filtrado mostrará radiactividad. El ARNt con valina se ha complementado al triplete GUU en el ribosoma y, por tanto, ha quedado atrapado por el filtro, lo que nos permite concluir que el triplete GUU codifica para la valina.

De esta manera, se utilizaba al ribosoma como lector de tripletes de secuencia conocida para comprobar qué ARNt se unía al triplete usado. Se trataba de un «método simple y directo que debería proporcionar una herramienta general para determinar la función genética de tripletes de secuencia conocida», en palabras del propio Nirenberg. Y realmente fue así. En 1966 todas las relaciones del código genético quedaron establecidas. La humanidad había sido capaz de descifrar el diccionario que permitía la comunicación entre las dos principales moléculas de la vida. Las posibilidades que ofrecía tal hazaña intelectual parecían inconmensurables.

El desciframiento del código sentó las bases para la futura lectura y reescritura de la información genética. La lectura del ADN culminó con el Proyecto Genoma Humano que, en el año 2001, publicó la primera secuencia completa de toda la información gené-

tica contenida en la célula de una persona. Nunca antes habíamos tenido la posibilidad de conocernos tan profundamente. La reescritura de la información genética, junto con el carácter universal del código, nos ha permitido desde introducir genes de una especie a otra para producir fármacos como la insulina a curar enfermedades de origen genético en humanos. Todos estos avances no hubieran sido posibles sin la aventura intelectual que tuvo lugar entre 1953 y 1966, y que ha conducido a la humanidad hacia nuevos horizontes jamás imaginados.

03

CAMBIOS EN LA INFORMACIÓN GENÉTICA: LAS MUTACIONES

La información del ADN es responsable, en buena parte, de cómo somos: es lo que nos hace humanos pero, a la vez, determina que seamos distintos entre nosotros. Estas diferencias se deben, en gran medida, a cambios en la información genética: las mutaciones.

Decía Alfred Nobel, el creador de los premios que llevan su apellido, que «la observación y la búsqueda de las similitudes y de las diferencias constituyen la base de todo el conocimiento humano». El desciframiento del código genético y la constatación de su universalidad determinaron la similitud que comparte toda la vida terrestre. Aunque la base genética es la misma, los organismos son distintos entre sí por las diferencias que encontramos en su ADN. Las bacterias se diferencian de las secuoyas por su información genética, no por la manera en que se lee esta información. Y estas diferencias no se dan solo entre especies: cada individuo es, genéticamente, distinto del resto de sus congéneres. La individualidad se halla en el ADN. Cada organismo, cada persona, se diferencia del resto por la información genética que hereda de sus progenitores, pero también por los errores propios que aparecen en su ADN. A estos errores se les conoce, de manera genérica, como *mutaciones*. Algunas de las mutaciones pueden darse sobre genes, es decir, sobre porciones del ADN que codifican directamente para proteínas. En estos casos, el código genético condiciona buena parte de

los efectos de los cambios del ADN sobre las proteínas, aunque el código en sí no muta.

Una mutación, entendida como cualquier cambio en el ADN, no es algo excepcional. Cada uno de nosotros acumula millones de mutaciones. Cada una de nuestras células es, estrictamente, una mutante, pues acumula cambios en el ADN que no estaban presenten en la célula de la que procede. Pero, a pesar de esta ubicuidad de los cambios, los seres vivos continúan funcionando. Esto es en parte debido a la acción de los mecanismos de reparación del ADN y en parte a que a veces las mutaciones pueden no afectar en nada al funcionamiento del organismo. ¿Qué mutaciones, pues, podemos encontrar en nuestro ADN?

UN MAR DE MUTACIONES

De los millones de mutaciones que acumulamos, algunas pueden afectar a unas pocas secuencias de nucleótidos (letras del ADN), mientras que otras pueden modificar fragmentos enormes del ADN. Los primeros casos suelen tener efectos menos evidentes que los segundos, aunque, en ocasiones, el cambio de una sola letra del ADN determina la aparición de una enfermedad o, incluso, puede hacer inviable la célula o al propio individuo. Los efectos de las mutaciones dependerán de diferentes factores como, por ejemplo, si es una mutación en la línea germinal que podrán heredar los descendientes, si altera una secuencia que regula únicamente la expresión de un gen, si produce cambios en alguna proteína, si el gen alterado es un regulador transcripcional, etc.

La primera consideración que hay que tener en cuenta es el carácter hereditario, o no, de una mutación. Aunque cada uno de nosotros lleve consigo un enorme número de mutaciones, solo unas pocas —las que aparezcan en las células *germinales*, las que darán lugar a las células reproductoras (óvulos y espermatozoides)— po-

drán pasar a la descendencia. El resto de mutaciones, denominadas *somáticas*, solo nos afectarán a nosotros y no las transmitiremos.

En este sentido, el cáncer se considera una enfermedad producida por la acumulación de mutaciones en diferentes genes que permiten a determinadas células crecer y dividirse de manera descontrolada e independiente del organismo. La mayoría de estas mutaciones aparecen en las células somáticas de la persona que lo padece. Sin embargo, algunas mutaciones pueden proceder ya de sus progenitores y predisponerlo a tener más probabilidades de desarrollar un cáncer en la vida adulta. Es el caso de las mutaciones hereditarias en los genes BRCA1 y BRCA2. Se ha demostrado que las personas que presentan ciertas mutaciones en estos genes tienen una mayor tendencia a desarrollar cáncer de mama que las que no portan estas alteraciones.

> Si no tuviéramos mutaciones genéticas, no seríamos nosotros mismos.
>
> STEVEN JOHNSON

Otro factor importante es el lugar del ADN en que aparece una mutación. Se estima que los genes codificantes (que expresan proteínas) representan un máximo del 3% de la secuencia total del ADN humano y, por tanto, parece más probable que una mutación aparezca en una secuencia no codificante que no se traduce a proteínas. Sin embargo, esto no significa que la mutación no tenga ningún efecto, ya que el 97% del ADN no codificante no está carente de sentido y puede participar, por ejemplo, en cuándo y cómo se transcribe un ARNm o en la formación de otros ARN como los ribosómicos o los de transferencia, afectando en última instancia a todas las proteínas. Estas mutaciones con efectos tan devastadores suelen ser raras, puesto que las células que las presentan no suelen sobrevivir.

Si la mutación aparece en un gen codificante, se trasladará al ARNm que se origina durante la transcripción. Es en este punto

en el que las mutaciones se relacionan con el código genético: solo las mutaciones que se encuentran en la secuencia de los genes que se transcribirán y acabarán formando parte de un ARNm maduro son traducidas; el resto de las secuencias no se traducirán y, por tanto, no serán «interpretadas» a través del código genético.

Mutaciones de pocas letras

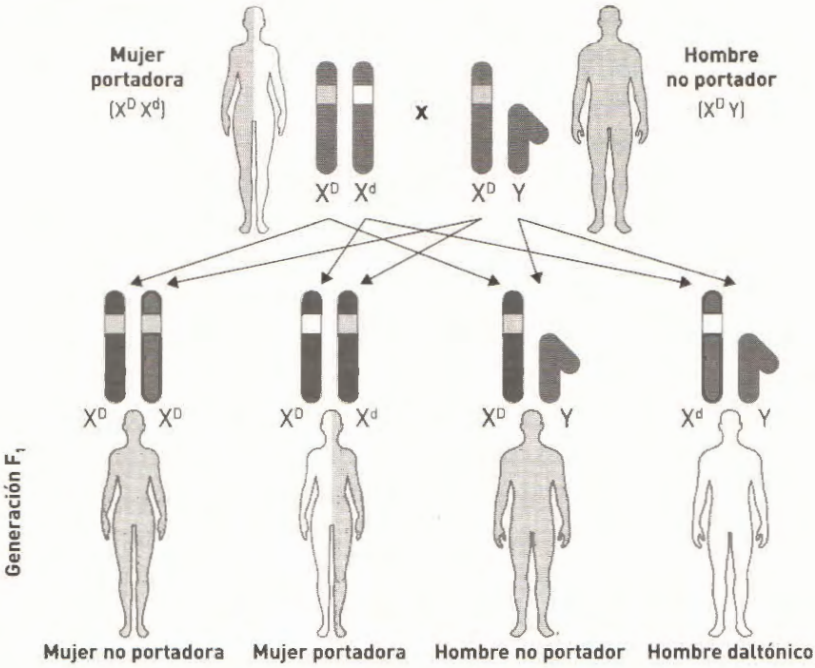
Una mutación puede consistir en el cambio de un único nucleótido (letra) —la sustitución de una base por otra, la modificación o la incorporación o pérdida de una base—. Como el código genético es redundante, algunos cambios de letra no tendrán efecto. Así, por ejemplo, si en una zona de la secuencia AUG GCC CUG UGG del ARNm de la insulina se cambia la sexta letra de una C a una U, se obtiene un triplete GCU que es sinónimo del original GCC, ya que codifican el mismo aminoácido, la alanina. La proteína tiene la misma secuencia pese al cambio de letra del ARN y, por tanto, ni la proteína ni su función se ven afectadas. Estas mutaciones son consideradas *mutaciones silenciosas*.

Algunos cambios de letra, sin embargo, producen un cambio de aminoácido, lo que se conoce como *mutación de cambio de sentido*. Así, por ejemplo, si en una secuencia de ARNm, al codón UGU, que codifica para la cisteína, le cambiamos la primera letra por una C, obtenemos el triplete CGU que codifica para la arginina. Este cambio de aminoácido es, precisamente, el que presentan, en una opsina —la proteína encargada de detectar los colores en nuestra retina—, algunas personas aquejadas de una forma concreta de ceguera selectiva al color.

En otras patologías, el cambio de letra puede tener un efecto más radical si provoca la aparición de un codón stop, lo que se conoce como *mutación sin sentido*. En estos casos, la proteína que se

> EL DALTONISMO, FRUTO DE UNA MUTACIÓN PUNTUAL

Un ejemplo de las consecuencias que tiene un simple cambio de letra en un codón es el daltonismo o ceguera selectiva al color. La más frecuente es la que afecta a la capacidad de discriminar el rojo o el verde. Esta alteración se hereda ligada al cromosoma sexual X, lo que explica su mayor frecuencia entre varones. Los hombres solo presentan un cromosoma X, si este posee la variante mutada (d) del gen, será daltónico. En cambio, como las mujeres presentan dos cromosomas X, necesitan poseer dos copias (dd) del gen para ser daltónicas. Con una sola copia (Dd), son portadoras, pero poseen una visión normal.



Esquema de la herencia de la ceguera selectiva al rojo o verde, ligada al cromosoma X. Los varones con una copia mutada del gen (d) ya son daltónicos; las mujeres necesitan dos copias (dd) para serlo.

genera es una versión truncada de la original, más corta y con un plegamiento completamente distinto. Es el caso, por ejemplo, de las mutaciones en el gen de la distrofina, que presentan algunas personas con distrofia muscular de Duchenne y que sufren una pérdida generalizada y progresiva de la masa muscular que les termina incapacitando. El gen mutado cambia un triplete para el aminoácido glutamina por un triplete de paro, un codón stop, durante la síntesis de la distrofina, de manera que la proteína normal, que está formada por más de 3500 aminoácidos, se transforma en una variante mutada con solo unos 1000 aminoácidos.

Finalmente, existen mutaciones que quitan (*deleciones*) o añaden (*inserciones*) nucleótidos, letras al ADN. Si se quitan o añaden 3 letras o cualquier número múltiplo de 3, se pueden obtener proteínas que presentan como único cambio la pérdida o incremento de uno o más aminoácidos. Estos aminoácidos de más o de menos pueden no afectar o pueden alterar profundamente la función de la proteína dependiendo de la función que los aminoácidos modificados tengan en la conformación estructural de la proteína. Así, por ejemplo, si a nuestra ya conocida secuencia AUG GCC CUG UGG... le quitamos las tres letras del segundo triplete (GCC) obtenemos la secuencia AUG CUG UGG... que codificará para una proteína de insulina idéntica a la normal a la que solo le faltará el segundo aminoácido, alanina, codificada por el GCC que se ha perdido.

Las inserciones o deleciones de un número no múltiplo de 3 (1, 2, 4, 5, 7, 8...) generalmente son mucho más drásticas porque cambian completamente la pauta de lectura del ribosoma y, de hecho, se las conoce como *mutaciones de cambio de la pauta de lectura*. Recordemos que la pauta de lectura por tripletes no solapantes es una de las características básicas del código genético. Si a la secuencia de la insulina que nos servía de ejemplo (AUG GCC CUG UGG... que codifica para metionina-alanina-leucina-triptófano-...) le quitamos una sola letra, en este caso la cuarta, obtene-

mos la secuencia AUG CCC UGU GGA... Esta secuencia codifica para una nueva proteína que solo conserva el primer aminoácido (metionina, codificado por el triplete AUG que no ha cambiado); el resto de aminoácidos son diferentes de los de la proteína original, ya que se han modificado todos los tripletes a partir de la mutación. Así, la nueva proteína estará formada por metionina-prolina-cisteína-glicina-..., completamente distinta de la secuencia normal anterior y, por tanto, con escasa capacidad para ejercer las funciones fisiológicas.

Mutaciones de grandes fragmentos

Las mutaciones que acabamos de ver consisten en modificaciones producidas por el cambio, inserción o deleción de unas pocas letras del ADN. Sin embargo, también se pueden producir mutaciones en las que participa un número mayor de nucleótidos y que suelen afectar a regiones enteras de los cromosomas, las largas cadenas de ADN que se encuentran en el núcleo. Estos cambios pueden alterar el funcionamiento de muchos genes y, por tanto, provocar efectos más espectaculares sobre el funcionamiento de la célula y del organismo que en las mutaciones anteriores.

Existen diferentes tipos de mutaciones que alteran regiones de los cromosomas (fig. 1). Así, por ejemplo, pueden darse pérdidas de regiones enteras (deleciones); regiones enteras duplicadas (duplicaciones) o colocadas al revés (inversiones) y regiones de dos cromosomas diferentes que intercambian su lugar (translocaciones).

Algunas de estas alteraciones cromosómicas, denominadas *estructurales*, producen cambios evidentes en el aspecto físico de las personas que las presentan, y que pueden afectar tanto a los rasgos morfológicos como a los cognitivos y psicológicos. Así, una dele-

Fig. 1

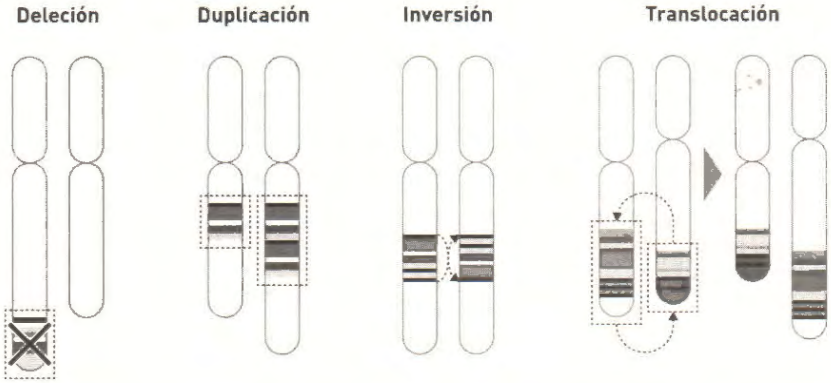


Imagen que ilustra algunas de las alteraciones de la estructura de los cromosomas.

ción de una porción del cromosoma 5 da lugar al síndrome del *cri du chat*, y una deleción de una porción del cromosoma 7 da lugar al síndrome de Williams.

Existen mutaciones, aún a mayor escala, que implican cambios en el número de cromosomas presentes en las células. Todas las células de un organismo, excepto los gametos, presentan el mismo número de cromosomas. En el caso de la especie humana, dentro de cada célula, exceptuando espermatozoides y óvulos, encontraremos 46 cromosomas; es decir, 23 parejas de cromosomas. Sin embargo, algunos individuos pueden presentar un número distinto de cromosomas, ya sea por la ausencia o por la presencia adicional de algún cromosoma. Los hombres con síndrome de Klinefelter presentan un cromosoma X adicional —XXY en vez de XY—; las mujeres con síndrome de Turner tienen un solo cromosoma X, en vez de la pareja XX que presentan la mayoría de las mujeres. La alteración en el número de cromosomas más conocida, sin embargo, es el síndrome de Down,

originado, entre otras causas, por la presencia de tres cromosomas 21.

EL ORIGEN DE LAS MUTACIONES

Las alteraciones en el número de cromosomas o en regiones dentro del cromosoma se producen por divisiones celulares defectuosas. Si este defecto se da durante la formación de las células reproductoras, la alteración la podrán presentar todos sus descendientes si esas células gaméticas terminan originando un cigoto. Durante la división normal para formar células reproductoras, denominada *meiosis*, primero se duplican todos los cromosomas —en el caso de los humanos, de 46 pasamos a 92—, y luego se forman cuatro células, cada una de ellas con 23 cromosomas. Si se produce algún fallo en esta redistribución, una de las células puede quedar con 22 cromosomas y otra con 24. La unión de estas células con otro gameto de 23 cromosomas, generará individuos con 45 o con 46 cromosomas.

Durante la meiosis, los cromosomas se alinean por pares e intercambian algunas regiones equivalentes (homólogas) mediante lo que se conoce como *recombinación meiótica* (fig. 2) De esta manera, se aumenta la variabilidad genética con una mayor diversidad de la descendencia. Si se producen errores durante esta recombinación, pueden dar lugar a alteraciones estructurales (translocaciones, inversiones, etc.).

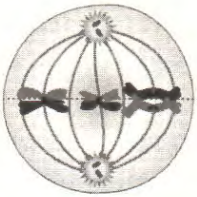
Las mutaciones cualitativamente menores como los cambios, inserciones o deleciones de pocos nucleótidos pueden presentar diferentes orígenes. Se pueden producir mutaciones espontáneas debidas al azar, a errores en los mecanismos de copia y reparación del ADN y por agentes externos, principalmente sustancias químicas o radiaciones, que, al actuar como mutágenos pueden alterar directamente la secuencia de nucleótidos de la cadena del ADN dando lugar a *mutaciones inducidas*.

Fig. 2

Profase I

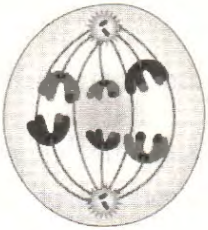
Los cromosomas se condensan, se aparean y se recombinan. Se rompe la membrana celular.

Metafase I



Las parejas de cromosomas homólogos se desplazan al ecuador de la célula.

Anafase I



Los cromosomas homólogos se desplazan a polos opuestos de la célula.

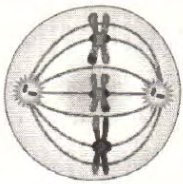
Telofase I y citocinesis

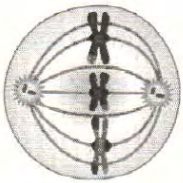
Se produce la división celular.

Profase II

Se inicia una nueva división celular.

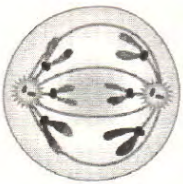
Metafase II

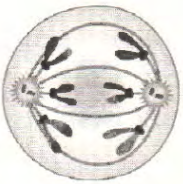




Los cromosomas (formados por dos cromátidas) se alinean en el ecuador celular.

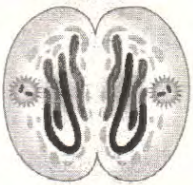
Anafase II

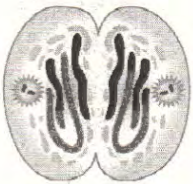




Las cromátidas de cada cromosoma se dirigen a polos opuestos de la célula.

Telofase II y citocinesis





Las células se dividen y se regeneran los núcleos. Se forman cuatro células finales.

Proceso de la meiosis en el que se aprecia la recombinación meiótica y su resultado (los cromosomas presentan diferentes sombreados indicando el origen de sus fragmentos).

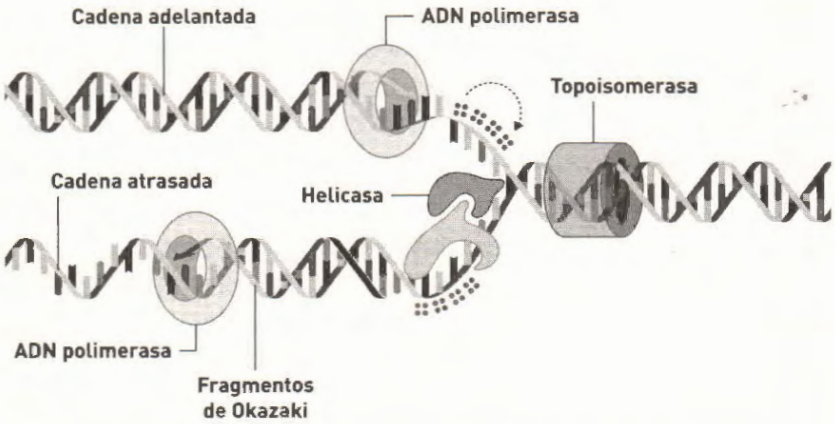
MUTACIONES ESPONTÁNEAS

Las mutaciones espontáneas se pueden producir durante la replicación del ADN; es decir, durante su duplicación, antes de que una célula se divida. Recordemos que el ADN forma una doble hélice, en la que las dos cadenas están fuertemente unidas entre sí por uniones químicas entre las bases de los nucleótidos de cada cadena denominadas *enlaces o puentes de hidrógeno*. Esta estructura condiciona la manera como se copia esta molécula. Cada cadena servirá de molde para la síntesis de una cadena complementaria. Pero, para poder utilizar el ADN como molde, la doble hélice debe abrirse dejando las dos cadenas separadas. Cada una de estas cadenas sirve de molde para formar nuevas dobles hélices complementarias, cada una con la misma secuencia de bases que la doble hélice inicial (parental). El proceso de replicación se inicia con una proteína denominada *helicasa*, cuya función es separar la doble cadena del ADN a modo de la apertura de una cremallera, lo que consigue mediante la rotura de los enlaces de hidrógeno que unen las bases de los nucleótidos (fig. 3).

> Las mutaciones genéticas siempre han ocurrido de manera natural.
>
> JAMES LOVELOCK

Cuando la molécula de ADN queda abierta, otra proteína, *la ADN polimerasa*, polimeriza la nueva cadena de ADN a partir de la incorporación de nucleótidos. La ADN polimerasa solo puede leer la cadena del ADN en un sentido y no en el opuesto. Como las dos cadenas que forman el ADN presentan sentidos de lectura opuestos, una de ellas puede duplicarse de manera continua, siguiendo el avance de la helicasa; la otra, sin embargo, debe leerse de manera discontinua ya que se sintetiza «a trozos», dando lugar a fragmentos nuevos de ADN de 100 a 200 nucleótidos que se conocen como *fragmentos de Okazaki*. Estos fragmentos, una vez sintetizados, se tienen que unir para dejar una cadena continua y

Fig. 3



En la cadena adelantada, la ADN polimerasa avanza en el mismo sentido que la helicasa. La ADN polimerasa de la cadena atrasada se desplaza en sentido contrario, generando los fragmentos de Okazaki.

sin interrupciones, lo que se consigue gracias a la acción de otra enzima, la *ADN ligasa*. A través de este mecanismo se producen dos cadenas esencialmente idénticas a las iniciales.

Las proteínas que se encargan de copiar el ADN no son infalibles y pueden llegar a cometer un error cada 100 000 nucleótidos copiados, lo que equivaldría a dos errores tipográficos en este libro. Puede parecer una buena media, pero si tenemos en cuenta que cada célula del cuerpo humano contiene 6 000 000 000 de bases en su ADN, cada vez que una célula se divide puede llegar a acumular unos 60 000 errores. Estos errores de replicación pueden ser de diferentes tipos. Por ejemplo, la maquinaria proteica puede saltarse unas pocas letras (se dice que «resbala» unas letras). Dependiendo de cómo se produzca este resbalón, se generan inserciones o deleciones de unas pocas letras. Otro error de replicación se produce por una lectura «defectuosa» de una letra del ADN, lo que condicio-

na que la ADN polimerasa introduzca una letra errónea en la cadena que está sintetizando.

Además de estas mutaciones relacionadas con errores en la replicación, se pueden producir lesiones espontáneas. La más frecuente es la desaminación, en la que una de las bases del ADN pierde parte de su estructura y se transforma en otra base distinta. Este cambio se produce sobre todo en las citosinas, aunque también puede darse en guaninas y adeninas. La desaminación de citosina produce uracilo, que normalmente se encuentra en los ARN. Cuando la ADN polimerasa lee el uracilo (U) en el ADN, coloca como nucleótido complementario una A (en vez de la G que correspondería a la C que se ha perdido).

Otra lesión espontánea, aunque poco frecuente, es la de la despurinización, en la que se pierde toda la base de uno de los nucleótidos derivados de la purina (una A o una G). En este caso, la hebra de ADN tiene una posición en la que le falta la base nitrogenada y, durante la replicación, se incorporará al azar cualquier nucleótido, lo que puede conducir a errores.

MUTACIONES INDUCIDAS

Las mutaciones pueden aparecer por la acción directa de agentes externos, los cuales producen alteraciones en el ADN que van desde pérdidas de una base a la rotura de una o ambas cadenas del ADN. El agente externo mutagénico al que estamos comúnmente más expuestos es la radiación ultravioleta, especialmente la proveniente de la luz solar.

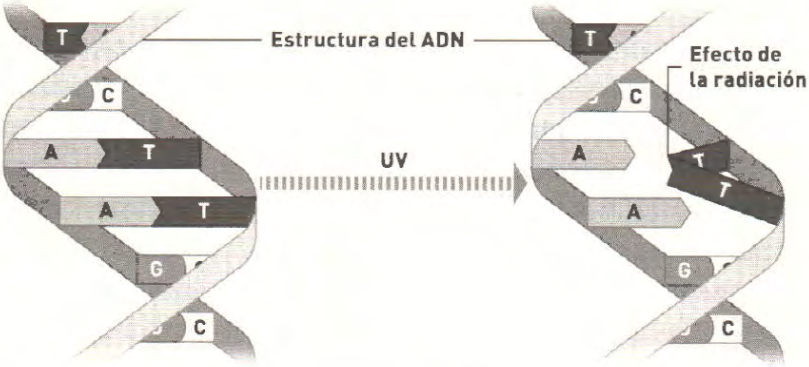
Radiaciones mutagénicas

La radiación ultravioleta se caracteriza por presentar una longitud de onda de entre 10 y 400 nm. La capa de ozono de nuestra atmós-

fera absorbe buena parte de la radiación ultravioleta solar dejando pasar solo la radiación ultravioleta A (rayos UVA, de 315 a 400 nm) y, muy parcialmente, la B (rayos UVB, de 280 a 315 nm). Estas longitudes de onda son capaces de interactuar con algunas moléculas orgánicas y producirles alteraciones. En el caso del ADN, los rayos ultravioletas pueden originar la unión química entre dos nucleótidos (C o T) que se encuentren adyacentes en la cadena del ADN. Se genera entonces un dímero que distorsiona la forma de la doble hélice e induce errores en la replicación del ADN. Aunque se pueden producir dímeros C-C, T-T, C-T o T-C, los más comunes suelen ser los dímeros de timina (T-T) (fig. 4).

Los rayos ultravioletas, por tanto, son un poderoso agente mutagénico para el que la evolución ha desarrollado su propio mecanismo de protección. Existen células especiales en nuestra piel que secretan y almacenan melanina, una sustancia capaz de absorber la luz ultravioleta. Cuando nos exponemos a una excesiva radiación ultravioleta, las células de la piel reconocen la agresión y li-

Fig. 4



Efecto de los rayos ultravioletas sobre el ADN. Se forma un dímero entre las bases de dos nucleótidos y se modifica la forma de la doble hélice.

beran y sintetizan más melanina. Como la melanina tiene un color marrón oscuro, la piel adquiere el típico tono de «bronceado». Si la radiación es excesiva, se produce un daño directo de las células de la piel, las cuales se inflaman y enrojecen con el característico color de la piel quemada por el sol. Las poblaciones humanas que se aclimataron a las regiones con mayor incidencia de sol, ecuatoriales y tropicales, presentan, de base, una mayor cantidad de melanina en la piel. Las poblaciones que habitan latitudes mayores, prácticamente no producen melanina, ya que la luz solar en la piel es necesaria para sintetizar la vitamina D y en esas latitudes la incidencia del sol resulta escasa para asegurar la producción necesaria de vitamina D. Hay que destacar que el efecto nocivo de los rayos ultravioletas es independiente de su fuente; es decir, tan peligrosos son los que recibimos del sol como a los que nos exponemos si, por ejemplo, vamos a un centro de bronceado. Los rayos UVA, provengan de donde provengan, son idénticos, indistinguibles, por lo que sus efectos sobre el ADN son los mismos.

Las mutaciones que esta radiación provoca contribuyen poderosamente a la aparición de melanomas, cánceres de piel que pueden resultar muy agresivos. Las cremas de protección solares cuentan con sustancias que absorben o bloquean los rayos ultravioletas y constituyen una herramienta imprescindible para disminuir la probabilidad de desarrollar cánceres cutáneos.

Otros tipos de radiaciones, con longitudes de onda más cortas que las de los rayos ultravioletas, también producen cambios en el ADN. Es el caso de los rayos X y los rayos gamma. Cada uno de ellos, además, puede penetrar más profundamente en los tejidos que los ultravioletas, los cuales afectan principalmente a la piel, porque solo penetran unos centímetros. Este poder de penetración es, de hecho, el que hace que se utilicen los rayos X para observar las estructuras internas de nuestro cuerpo y su poder mutagénico es el que hace recomendable que las embarazadas no se sometan a esta radiación para que el feto no sufra alteraciones

mutagénicas. Aunque los rayos X pueden producir, en algunas circunstancias, cambios en las letras del ADN, su efecto más significativo (al igual que ocurre con los rayos gamma) es la rotura de una o de las dos cadenas del ADN, hecho que puede resultar letal sobre la célula que lo sufre.

Sustancias mutagénicas

Además de las radiaciones, diferentes compuestos químicos pueden actuar sobre el ADN y producir mutaciones. Así, por ejemplo, el ácido nitroso (HNO_2) puede producir desaminaciones, como las espontáneas y, por tanto, convertir algunas citosinas en uracilos. Otros compuestos, como la nicotina, son capaces de añadir pequeñas moléculas a los nucleótidos cambiando su interacción con otros nucleótidos, lo que provoca que las proteínas que replican el ADN cometan errores y se produzca la unión física entre nucleótidos o, incluso, la rotura de la cadena de ADN.

Otros compuestos capaces de alterar el ADN son las especies reactivas del oxígeno, pequeñas moléculas o átomos producidos por diferentes procesos dentro de la célula, que oxidan los ácidos nucleicos, especialmente la guanina. Cuando se produce un exceso de estas especies se habla de *estrés oxidativo*, que se puede producir en respuesta a diferentes procesos fisiológicos, como, precisamente, el cáncer. En un ejemplo de retroalimentación positiva, las especies reactivas de oxígeno producen mutaciones que aumentan la probabilidad de desarrollar cáncer, lo que, a su vez, supone un incremento de las especies reactivas de oxígeno y, por tanto, de mutaciones, lo que acelera la progresión del proceso tumoral. Las células disponen de mecanismos moleculares para controlar la producción de las especies reactivas de oxígeno y de sus efectos, además de toda una batería de proteínas implicadas en diferentes procesos de reparación del ADN.

Agentes genéticos mutagénicos

Además de los compuestos químicos y de las radiaciones, las mutaciones inducidas también pueden producirse por agentes genéticos, como, por ejemplo, los virus o elementos de nuestro propio ADN denominados *transposones* o *elementos genéticos móviles*. Estos últimos elementos son secuencias de ADN que poseen la capacidad intrínseca de cambiar de posición dentro del genoma. Fueron descubiertos por la genetista estadounidense Barbara McClintock, incluso antes de que se conociese la estructura del ADN.

Cuando estos elementos cambian de posición dentro del ADN, producen mutaciones tanto en las localizaciones donde se encontraban como en su nuevo destino. Si el elemento transponible se inserta en el interior de un gen se produce una inserción de una gran cantidad de bases que harán que el gen pierda su función. Si el transposón vuelve a saltar, el gen recupera su funcionalidad y vuelve a expresarse. Se calcula que en el genoma humano hay un 44% de elementos transponibles, porcentaje que puede ascender hasta el 66% teniendo en cuenta fragmentos repetidos y secuencias cortas derivadas de ellos. Se estima que tenemos más de 1000 genes regulados, directa o indirectamente, por secuencias procedentes de elementos transponibles.

Hasta el momento se conocen dos tipos de elementos transponibles: los de clase I o retrotransposones y los de clase II o transposones de ADN. Los retrotransposones cambian de sitio a través de la transcripción inversa o retrotranscripción. La transcripción inversa característica de los transposones de clase I es similar al proceso de transcripción que conocemos, pero con la diferencia de que se produce en sentido inverso. Es decir, si en el proceso clásico de transcripción se obtiene una cadena simple de ARN a partir de una doble cadena de ADN, en la transcripción inversa se obtiene una molécula de ADN a partir de una molécula de ARN. Los retrotransposones representan el 90% de los elementos transponibles

de nuestro genoma. El 10 % restante corresponde a los transposones de clase II, los cuales saltan dentro del ADN sin necesidad de pasar por un estadio intermedio de ARN y, por tanto, de manera independiente de la retrotranscripción.

La retrotranscripción es también propia de algunos tipos de virus, los retrovirus, cuya información genética se encuentra almacenada no en moléculas de ADN sino en moléculas de ARN, como es el caso del virus VIH (causante del sida), el HTLV-1 (capaz de causar tumores en células del sistema inmunitario) o los virus de la gripe, el del ébola, la polio o el sarampión. Al infectar una célula, los virus de ARN retrotranscriben ADN a partir de su ARN e integran el ADN en el genoma de la célula infectada (fig. 5).

El otro gran grupo de virus, los virus de ADN o adenovirus, transportan su información genética en moléculas de ADN. El virus del papiloma humano, el de la hepatitis B o el del herpes son ejemplos de este tipo de virus, los cuales no necesitan realizar ninguna retrotranscripción para insertar su ADN en el genoma de la célula infectada.

La inserción del ADN vírico puede producir cambios sobre los genes de la célula, análogos a los de los transposones. Si unos u otros se insertan en alguna región del ADN que produce la codificación de una proteína, esta proteína dejará de ser funcional. Si se insertan en las regiones que controlan la expresión de una proteína, la proteína se desregulará. Ambas alternativas pueden ser catastróficas para la viabilidad de la célula.

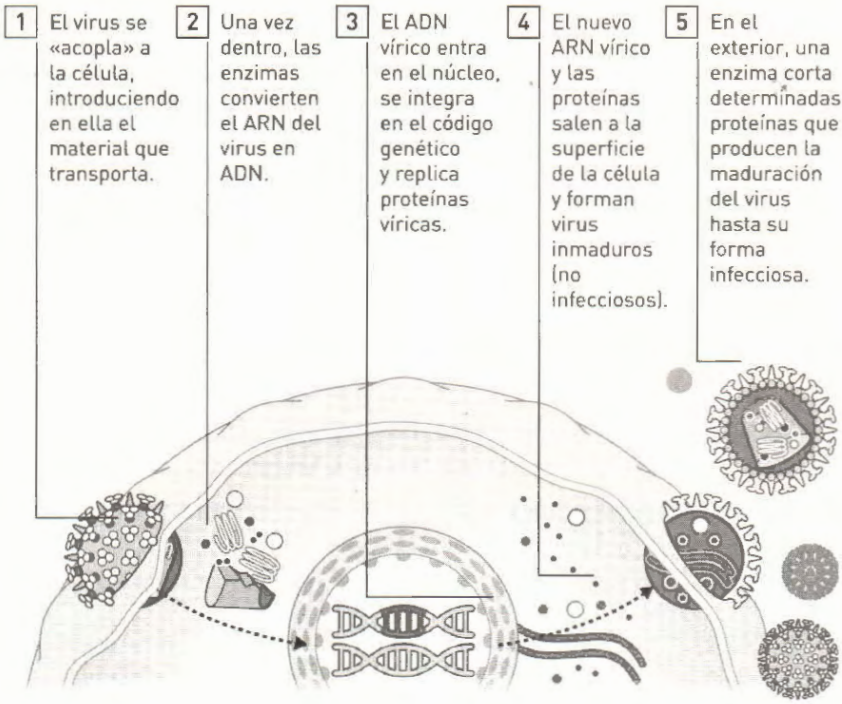
Las similitudes entre los transposones y los virus son enormes y, de hecho, se podría considerar que los virus son elementos transponibles con la capacidad de saltar de célula en célula o, alternativamente, se podría definir a los transposones como virus que han perdido su capacidad de salir de la célula en la que se alojan. Sea como sea, ambos han sido reconocidos como una fuerza a tener en cuenta en el estudio de la evolución de las especies por su capacidad de generar cambios drásticos que contribuyen a la

> LOS GENES SALTARINES DEL MAÍZ

A finales de la década de 1920, la científica estadounidense Barbara McClintock estudiaba diferentes caracteres en plantas de maíz cuando observó que una misma mazorca podía presentar granos de distintos colores. Este hecho la sorprendió, ya que todos los granos contenían la misma información genética: la que habían heredado de sus padres. McClintock dedujo que debían existir unos elementos móviles que saltaban dentro del genoma de los granos del maíz, alterando con sus cambios diferentes genes y produciendo, así, la variedad de colores que observaba. Estas secuencias genéticas que cambiaban de posición eran los elementos transponibles. Su descubrimiento, publicado a finales de la década de 1940, le valió en 1983 el premio Nobel.

— Barbara McClintock trabajando con maíz en su laboratorio de Cold Spring Harbor, Nueva York, en abril de 1963.

Fig. 5



Esquema del ciclo vital de un virus de ARN, el VIH, causante del sida. Se aprecia la inserción del ADN vírico (sombreado) obtenido a partir del ARN por retrotranscripción.

variabilidad genética, incrementando la diversidad dentro de las especies y entre ellas. Sin embargo, no hay que obviar que estos elementos pueden ser los causantes de alteraciones genéticas que desestabilicen las células y que acaben provocando tumores o la propia muerte celular.

Como acabamos de ver, existen diferentes mecanismos que pueden modificar el ADN de una célula. Aunque algunos de estos procesos son más probables y constantes que otros, sobre el ADN

original siempre se producen modificaciones, constituyendo errores en muchos casos. La evolución ha dotado a las células de mecanismos de reparación y corrección de esos errores que tratan de minimizar la cantidad de alteraciones que acaban persistiendo y que tratan de asegurar la fidelidad con respecto a la información del ADN original.

LA REPARACIÓN DEL ADN

De la misma manera que existe el código PAS (Proteger, Avisar y Socorrer) para realizar los primeros auxilios, en los casos de accidentes, la célula también tiene su propio código PAS frente a un daño, destinado a Percibirlo, Actuar para repararlo y Sobrevivir para continuar con sus funciones. Existen diferentes mecanismos de reparación del ADN, algunos de ellos especializados en la corrección de la alteración de la secuencia de nucleótidos, y otros dedicados a la reparación de daños mayores, como la rotura de las cadenas.

Uno de los mecanismos especializados en la corrección de las alteraciones de la secuencia de nucleótidos se encarga de reparar modificaciones químicas que aparecen en algunos nucleótidos, pero que no alteran la forma de la doble hélice. A este mecanismo se le conoce como *reparación por escisión de bases* (la porción de los nucleótidos que determina qué letra es). En estos casos, se corta y se desprende la base dañada, generándose un espacio en blanco que debe ser nuevamente ocupado. Para ello se corta un fragmento del ADN que lo incluya y que será rellenado por la maquinaria encargada de la replicación y unido al resto de la cadena por las proteínas ligasas.

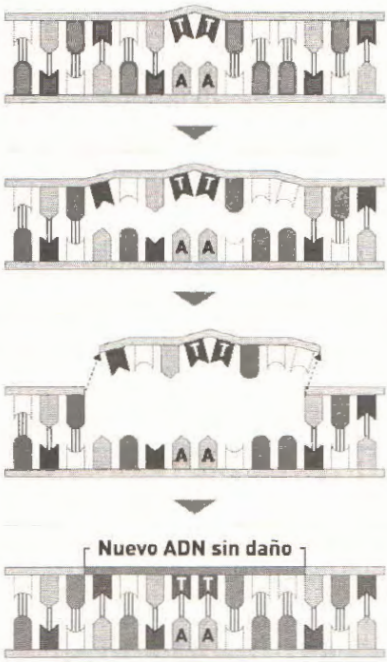
Existe otro mecanismo que interviene en las modificaciones químicas que alteran la forma de la doble hélice, como resultado, por ejemplo, de la acción de algunos mutágenos químicos o, principalmente, de la radiación ultravioleta. El mecanismo que

repara estas mutaciones es la *reparación por escisión de nucleótidos* y puede darse durante la propia replicación o de manera general. Si se da durante la replicación, es la propia maquinaria de la replicación la que detecta el daño producido; en el caso de la reparación general, existen proteínas específicas que detectan y señalan el daño. En cualquiera de los dos casos la célula procede como en el mecanismo anterior: la secuencia del ADN que contiene el error se corta y la maquinaria de transcripción rellena el hueco que se ha generado (fig. 6).

En este mecanismo de reparación participan diferentes proteínas. Si alguna de ella está mutada y no puede ejercer su función de manera eficaz, se acumulan este tipo de errores, mutaciones en el ADN, lo que aumenta el riesgo de desarrollar procesos tumorales y otras afecciones. Es el caso, por ejemplo, de los afectados por una genodermatosis, *Xenoderma pigmentosum*, enfermedad causada por mutaciones en ocho genes implicados en la reparación del ADN, de los que siete lo están en la reparación por escisión de nucleótidos. El resultado de las mutaciones en estos genes provocan una hipersensibilidad a la luz y una alta incidencia de cáncer cutáneo. Otras enfermedades producidas por mutaciones en genes que regulan los mecanismos de reparación son la tricotiodistrofia y el síndrome de Cockayne, que se caracterizan por presentar fotosensibilidad y deficiencia intelectual.

Un tercer mecanismo de reparación es el que actúa sobre los nucleótidos mal apareados; es decir, detecta, por ejemplo, parejas A-G o T-C. Este mecanismo se conoce como *reparación de los desapareamientos* y presenta un funcionamiento similar a los anteriores. Una proteína detecta y se une específicamente a los nucleótidos mal apareados facilitando la degradación de un fragmento corto de la cadena de ADN que contenga los nucleótidos mal apareados. Quedará entonces un vacío que será rellenado por la acción de la ADN polimerasa. El mecanismo de reparación de los desapareamientos suele actuar sobre errores producidos por la maquinaria

Fig. 6



La maquinaria de reparación detecta un daño que altera la forma de la doble hélice (un dímero de timina en este caso).

Mediante la rotura de los enlaces de hidrógeno de las bases adyacentes, se abre la doble hélice.

Se desprende un fragmento corto de ADN que contiene el dímero de timina.

La ADN polimerasa replica la cadena a partir de la complementaria. Las ligasas unen el nuevo fragmento a la cadena de ADN original.

Esquema de la reparación por escisión de nucleótidos. En el esquema no se representan las proteínas involucradas en la reparación.

de replicación, que, aunque cuentan con un sistema de comprobación de errores propio, no es totalmente infalible.

Si alguna de las proteínas de este sistema de reparación está alterada, las consecuencias sobre el individuo que presenta la mutación podrá desencadenar diversas patologías. Sirve de ejemplo el síndrome de Lynch o cáncer colorrectal hereditario no poliposo, una de las enfermedades hereditarias más comunes, causante del 15 % de todos los casos de cáncer de colon. En estos pacientes, como el mecanismo de reparación de desapareamientos no funciona correctamente, se acumulan las mutaciones a una mayor tasa que en

aquellas personas cuyo mecanismo funciona correctamente. Esta acumulación de mutaciones hace más probable que se alteren proteínas que participan en la estabilidad celular y que, por tanto, se pueda desarrollar un proceso cancerígeno.

Los mecanismos que acabamos de indicar actúan sobre cambios relativamente menores del ADN. Sin embargo, hemos visto que determinados agentes, como los rayos X, son capaces de dañar con mayor severidad esta molécula, concretamente produciendo cortes en las dos cadenas del ADN. Para estos casos, la célula presenta tres posibles mecanismos de actuación.

El primero de ellos, denominado *unión de los extremos no homólogos*, se basa en la acción de un tipo especial de ligasa, las proteínas encargadas de unir cadenas de ADN. Se trata de una intervención rápida basada en la detección de fragmentos similares, muy pequeños, entre las cadenas que se quieren reparar. Este sistema de reparación no es muy preciso y puede generar deleciones e inserciones si los fragmentos reconocidos, aunque similares, no fuesen exactamente los que se corresponden con esa zona del ADN.

El segundo mecanismo, la unión de extremos mediada por microhomologías, se basa en la detección de regiones microhomólogas, es decir, fragmentos de una cadena parecidos a los de la otra cadena que se utilizan como guías para conducir la unión de las cadenas separadas. Por su manera de funcionar, este mecanismo siempre genera deleciones, por lo que, a pesar de que esté arreglando un daño del ADN, está introduciendo mutaciones en la secuencia.

El tercer mecanismo, la recombinación homóloga, es más preciso, aunque también más lento. En este mecanismo se utiliza una secuencia idéntica o muy parecida para reparar la rotura producida. Esta secuencia la encuentra, por ejemplo, en el cromosoma homólogo. Recordemos que en cada una de nuestras células tenemos 23 parejas de cromosomas, 46 cromosomas en total. Pues bien, si se rompen las cadenas, supongamos, de uno de los cromosomas 2, este mecanismo utiliza el otro cromosoma 2 como molde para repa-

rar la rotura del primero. Este mecanismo es muy parecido al de la recombinación meiótica al que nos hemos referido anteriormente. La disminución de las tasas de recombinación homóloga causa la ineficiente reparación del ADN que puede conducir a la aparición del cáncer. Es el caso de los genes BRCA que son supresores de tumores, cuyo mal funcionamiento se relaciona con un incremento del riesgo de cáncer de mama y de ovario.

La recombinación homóloga es también clave para el uso de la tecnología CRISPR/Cas, mediante la que se puede modificar, con altísima precisión, el ADN de los organismos. Ya se han realizado las primeras pruebas con humanos y los resultados son esperanzadores, como veremos en el siguiente capítulo.

04

MÁS ALLÁ DEL CÓDIGO: LA MODIFICACIÓN DE GENES

El descubrimiento del código genético ha abierto todo un abanico de posibilidades en el campo del tratamiento de enfermedades y patologías de origen genético. Nos hallamos en un momento en el que es posible modificar el ADN y editar los genes con un nivel de precisión nunca antes imaginado.

Tal y como hemos visto en los capítulos anteriores, el código escrito en nuestros genes puede sufrir modificaciones que, tras ser interpretadas, creen esas proteínas aberrantes que están detrás de las mutaciones causantes de los procesos tumorales, así como de los problemas musculares, metabólicos, de crecimiento... Gracias al descifrado del ADN, la ciencia ha conseguido identificar tales mutaciones, paso previo para actuar en consecuencia y desarrollar programas de prevención y curación.

Hoy en día, secuenciar y estudiar el ADN ya es una actividad casi rutinaria. Cada uno de nosotros puede tener secuenciado su genoma en menos de 24 horas y, con ello, saber qué tipo de alteraciones se presentan en nuestros genes y cómo eso puede llegar a afectar a nuestra salud. Existen más de 10 000 enfermedades genéticas incluidas en la información que se encuentra en las células de los seres humanos; muchas de ellas se manifiestan desde el nacimiento, otras lo hacen en momentos posteriores, y otras aumentan la predisposición a padecer un determinado tipo de dolencia. Actualmente ya es posible la identificación de mutaciones en el ADN causante de

diversas patologías, lo que puede ser muy útil a la hora de asesorar a pacientes con antecedentes de enfermedades familiares, ya que esas mutaciones pueden haber sido heredadas y estar impresas en su propio ADN. Conocerlo a tiempo nos dará las claves para prevenir su aparición o favorecer un diagnóstico temprano.

Pero las posibilidades que brinda haber descifrado el código genético no se limitan solo a la lectura y diagnóstico de las enfermedades genéticas, sino que van más allá al permitir editar la molécula de la vida, esto es, crear nuevos segmentos de ADN o modificar el nuestro propio para contribuir a sanar enfermedades. Se han generado en laboratorio microorganismos cuyo genoma se ha alterado de tal manera que pueden producir proteínas humanas tales como la insulina o la hormona de crecimiento recombinante, de las cuales hablaremos con más detalle en este mismo capítulo. Incluso se ha logrado algo que promete grandes logros en la investigación del futuro. Nos referimos a la edición de los propios genes contenidos en diversos tejidos de un ser humano. Actualmente, tenemos en nuestras manos la posibilidad de erradicar las enfermedades genéticas gracias a la nueva técnica de edición de genes, el CRISPR. Esta técnica ha abierto un amplísimo campo de investigación que nos permite eliminar o sustituir secuencias de ADN mutadas por sus homólogas sanas. Ya existen resultados esperanzadores con animales de laboratorio en los que enfermedades como la distrofia muscular de Duchenne o determinados tipos de cáncer han tenido un pronóstico favorable.

Pero no todo acaba aquí. ¿Y si fuésemos capaces de cambiar el código genético, hasta ahora universal, para introducir en él nuevos caracteres que la naturaleza no ha seleccionado? Ya se han producido los primeros estudios *in vitro* con resultados positivos que de generalizarse permitirían fabricar proteínas con nuevos componentes.

Sin duda, la puerta está abierta a nuevos tratamientos impensables hasta hace unos años, desde las técnicas gracias a las cuales podemos encontrar mutaciones en nuestros genes hasta las que

nos permiten solucionarlas. En el horizonte comienzan a vislumbrarse nuevas terapias, retos y esperanzas.

PREVENIR DESDE LOS GENES

Como hemos dicho, cada día, a cada momento, estamos expuestos a agentes mutágenos que provocan cambios y alteraciones en nuestro ADN. Estos cambios pueden llegar a generar mutaciones perjudiciales que alteran la información contenida en nuestros genes y que por ello nos pueden hacer más susceptibles a padecer una determinada enfermedad. La ciencia ya está preparada para detectar estas mutaciones, gracias a numerosas técnicas entre las que se incluye la secuenciación directa de nuestros genes, para así poner remedio o favorecer la prevención, especialmente en enfermedades como el cáncer. Sin embargo, no todas las mutaciones en las regiones codificantes de nuestros genes generan problemas a la hora de traducirse en proteínas.

> La secuenciación del genoma proveerá la explicación última de lo que es un ser humano.
>
> WALTER GILBERT

Cuando una célula identifica que su ADN ha incorporado una mutación en una región esencial que puede llegar a ser perjudicial, se pone en marcha un complejo programa de suicidio celular llamado *apoptosis*, que acaba con la vida de la célula antes de que potencialmente pueda convertirse en una célula maligna. Pero en otras ocasiones este proceso de suicidio celular no se activa correctamente y la célula acumula mutaciones en su genoma. En este caso, como comentamos en el capítulo anterior, si la mutación afecta a la línea germinal es posible pasarla a la descendencia.

Conocemos dos casos especialmente relevantes de tumores hereditarios o familiares: el cáncer de colon y el de mama.

Determinados tipos de cáncer colorrectal son causados por el llamado *síndrome de Lynch*, conocido como cáncer colorrectal hereditario y que presenta mutaciones en los genes MSH2, MLH1, MSH6 y hPMS2, los cuales codifican proteínas implicadas en la reparación del ADN. Cerca del 3% de los casos de cáncer colorrectal se deben a este síndrome, y las personas que lo padecen tienen muchas más probabilidades de presentar cáncer colorrectal o de endometrio (en el caso de las mujeres), especialmente a una edad más temprana, antes de los cincuenta. Las personas con síndrome de Lynch también tienen mayores probabilidades de presentar otros cánceres, incluido el cáncer de ovario, estómago, hígado, riñón, cerebro y piel.

También se ha detectado que las mutaciones en los genes BRCA1 y BRCA2, ya citados anteriormente, están relacionadas con cánceres hereditarios, en este caso de mama y ovario. Han sido numerosos los estudios encaminados a encontrar mutaciones en estos genes y, con ello, poder proporcionar consejo genético, ya que sabiendo el lugar exacto donde se localiza la mutación en una determinada familia, será mucho más fácil identificar su presencia en familiares que aún no han desarrollado el proceso tumoral.

Tanto BRCA1 como BRCA2 pueden presentar numerosos tipos de mutaciones que afectarán directamente a la interpretación que nuestras células hacen de la secuencia de tripletes siguiendo el código genético. La mayoría de esas mutaciones provocan cambios en la pauta de lectura debido a inserciones o deleciones de nucleótidos. Un ejemplo es el caso de la mutación del codón 23 del exón 2 de BRCA1, en el que se observa con frecuencia una deleción de un par de bases contiguas, el par AG. Dado que esto ocurre al inicio de la lectura, la eliminación de estos dos nucleótidos genera una proteína aberrante. El hecho de que el complejo BRCA no sea funcional ocasiona que en el ADN se acumulen roturas y otras alteraciones que pueden desembocar en un proceso tumoral.

Como se ha indicado, estas mutaciones ya se pueden detectar secuenciando los genes. Desde que en 1980 el bioquímico británico

Frederick Sanger desarrollase el primer método de secuenciación basado en la fragmentación de las largas cadenas de aminoácidos presentes en las moléculas de proteínas para después recomponer la estructura completa de forma manual, la rápida mejora de las técnicas de secuenciación en el ámbito de la medicina molecular ha hecho avanzar de manera espectacular la calidad del diagnóstico genético. Para cada tipo de enfermedad existe un grupo de genes susceptibles de presentar mutaciones que son los que se secuencian en cada caso, reduciendo así la necesidad de una secuenciación de todo el genoma.

Con la llegada del siglo XXI, entró pisando fuerte en el campo de la biotecnología una técnica de secuenciación del ADN que está acelerando y abaratando la búsqueda poblacional de mutaciones. Se denomina *next generation sequence* (NGS) y permite secuenciar un genoma completo en menos de 24 horas con una altísima fiabilidad en la lectura. Con ello, podemos tener miles de genomas secuenciados que es posible comparar para establecer correlaciones entre mutaciones en el ADN y patrones de salud o enfermedad. Esto abre un nuevo campo en la medicina preventiva, ya que conociendo el genoma de una persona y analizando sus variantes génicas y sus mutaciones, podremos establecer mapas de posibles enfermedades o patologías que nuestro ADN señale como más probables, pudiendo actuar en consecuencia con una medicina preventiva clara. Para ello se analizan determinadas variantes en nuestro genoma que pueden darnos claves de nuestra condición física y metabólica. Sin ir más lejos, conocemos que el gen ACTN3 puede estar implicado en la resistencia física. En este caso, la pérdida de función de este gen genera una ventaja, ya que en ausencia de esta proteína las fibras musculares presentan mayor resistencia aeróbica, si bien la velocidad se ve disminuida. De hecho, los deportistas de élite presentan en un elevado porcentaje este gen mutado, con su consiguiente pérdida de función. La mutación más común es la que presenta el triplete 577, en el que una simple mutación de

cambio de base, en la que una C es sustituida por una T, hace que el aminoácido arginina (CGA) se sustituya por un codón stop (UGA), lo que genera una proteína incompleta, ya que finaliza prematuramente su síntesis.

El descifrado de nuestro código genético ha conseguido que conocer información como esta, con un simple test genético, sea ya una realidad.

ENSAMBLAR NUEVO ADN

Gracias al profundo conocimiento que la ciencia nos ha brindado en cuestión de codificación en nuestro ADN, hemos sido capaces no solo de leer el genoma a través de su cifrado en tripletes, sino también de modificarlo y editarlo a voluntad. A lo largo de la historia, a medida que aumentaba el conocimiento sobre lo que nuestro ADN dicta, también era mayor la urgencia de aprender a modificar esas órdenes. En los inicios de la genética moderna los primeros organismos modificados genéticamente fueron productos agrícolas, con el fin de mejorar sus características; pero esta modificación era al azar. Durante la historia hemos sido eficientes realizando estas selecciones, pero sin entender cómo funcionaba el proceso en su totalidad. De hecho, en la década de 1960, cuando ya se sabía que la radiación provocaba mutaciones en el ADN, era común en los laboratorios el bombardeo indiscriminado de semillas para crear mutaciones al azar y así cambiar, también al azar, las propiedades de esas plantas. Se generaban mutaciones (roturas, mutaciones puntuales, deleciones...) en los genes de las semillas, creándose así algunas variedades, como las de la mazorca de maíz, que hoy siguen comercializándose. Este bombardeo generaba cambios en los nucleótidos que daban lugar a nuevos codones o nuevas pautas de lectura que, tras ser leídas en sus tripletes correspondientes, generaban nuevas proteínas. En la mayoría de las ocasiones estos

> SANGER, PADRE DEL MÉTODO DE SECUENCIACIÓN

— El bioquímico inglés Frederick Sanger en un retrato de 1973.

En 1980, el bioquímico británico Frederick Sanger obtuvo el premio Nobel de Química por sus contribuciones acerca de la determinación de secuencias de bases en ácidos nucleicos. No era su primer Nobel: en 1958, y en la misma categoría, recibió otro por su trabajo sobre la estructura de las proteínas, en especial la insulina. Todo un logro, pues desde que estos premios se instituyeron en 1895, solo cuatro personas han sido distinguidas con dos. Nacido en Gloucestershire en 1918, Sanger tuvo como primera vocación la medicina. Todo cambió durante su etapa de estudios en el St John's College de Cambridge, donde tuvo la oportunidad de asistir a las clases de algunos de los mejores bioquímicos de la época, como sir Frederick Gowland Hopkins o Malcolm Dixon. En 1943, el mismo año en que se doctoró, empezó a trabajar en el Departamento de Bioquímica de Cambridge. Fue en él donde sus investigaciones fructificaron en el descubrimiento de la secuencia de las proteínas, paso previo a una de sus aportaciones más importantes: el descubrimiento de la secuencia de la insulina. Igualmente, Sanger contribuyó a determinar la secuencia de nucleótidos del ADN. El método que ideó a tal fin y con el que consiguió la secuenciación completa del virus bacteriófago Φ-X174 se basaba en incorporar en la síntesis de la hebra de ADN nucleótidos con marcas de fósforo radiactivo que se podían leer en una placa de radiografía. En pocos años, la técnica evolucionó para incorporar marcas fluorescentes legibles con un láser. Estos descubrimientos revolucionaron la manera de entender el ADN y las enfermedades asociadas, y posibilitaron el arranque del proyecto Genoma Humano.

cambios no producían organismos viables o, de producirlos, no eran más aptos que los primeros.

Esta mutagénesis al azar fue pronto sustituida por un proceso en el que se puede dirigir el tipo de mutación deseada en una proteína mediante la manipulación del ADN. Para ello se usan pequeños fragmentos de ADN artificial sintetizados en laboratorio y en los cuales se incluye la mutación que queremos perpetuar en el ADN. Estas mutaciones suelen ser pequeños cambios de nucleótidos que generan cambios en los codones codificantes, lo cual da lugar a cambios de aminoácidos que varían la secuencia de la proteína. No solo esto, sino que hemos sido también capaces de introducir genes en el interior de organismos de otra especie, generando así organismos transgénicos.

Y es que el conocimiento del lenguaje del ADN, el código genético, es el primer paso para desarrollar una tecnología con la cual sea posible cortar y pegar genes, incluso editar el texto contenido en nuestro genoma. La tecnología capaz de realizar estas modificaciones es la ingeniería genética, y en las últimas décadas ha favorecido el avance de la investigación aplicada a la sanidad humana y a la cura de enfermedades. Por otra parte, la utilización de proteínas recombinantes como productos terapéuticos, vacunas y herramientas en ensayos de diagnóstico se extiende cada día más.

La tecnología del ADN recombinante

El hecho de que el código genético sea universal nos ofrece un sinfín de posibilidades, ya que si insertamos un fragmento de ADN de una célula de otra especie, este se expresará, es decir, transformará la información en las proteínas necesarias para su funcionamiento y desarrollo. En la década de 1970 se comenzaron a realizar experimentos que consistían en insertar pequeños plásmidos (diminutos fragmentos circulares de ADN) en el interior de las bacterias,

los cuales les conferían propiedades muy concretas, fundamentalmente de resistencia a antibióticos. Fue entonces cuando se demostró que transferir un gen de un organismo a otro era posible. Hoy en día, gran parte de la industria biotecnológica se basa en la producción de microorganismos recombinantes capaces de producir elevadas cantidades de proteínas. Es la llamada *tecnología del ADN recombinante*. El procedimiento es teóricamente sencillo: se usan células vivas como pequeñas fábricas de proteínas en miniatura, mediante la transferencia de genes en su interior.

Así, la tecnología de ADN recombinante requiere introducir el gen de interés en un sistema que propicie la replicación y transcripción del gen dentro de una célula viva, por lo general una bacteria no patógena. En este caso, el gen debe unirse a un vehículo que nos permita introducirlo en dicha bacteria y permitir su expresión. Para ello, el gen se selecciona mediante el corte con enzimas de restricción, se purifica y se une al vehículo. Estos vehículos suelen ser plásmidos, fragmentos relativamente pequeños de ADN circular que contienen las secuencias necesarias para la replicación y la transcripción del gen. Así, el gen se une mediante enzimas ligasas al plásmido cortado previamente con las mismas enzimas de restricción. El tipo de plásmido que se utilizará para clonación molecular depende en la elección del organismo huésped, el tamaño de ADN que será clonado y de la manera en la que el ADN foráneo será expresado. Una vez generado el plásmido con el gen de interés, se introduce en una bacteria que se dividirá, copiando en cada división el plásmido y dando lugar a grandes cantidades de ADN y proteína (fig. 1). Todo este proceso es posible ya que la bacteria es capaz de leer el ADN del mismo modo que lo hace una célula humana, utilizando el código genético universal.

Quizás uno de los ejemplos más relevantes de la tecnología del ADN recombinante, por su amplio uso a nivel mundial, sea la ya citada de la insulina recombinante, inyectada a millones de pacien-

Fig. 1

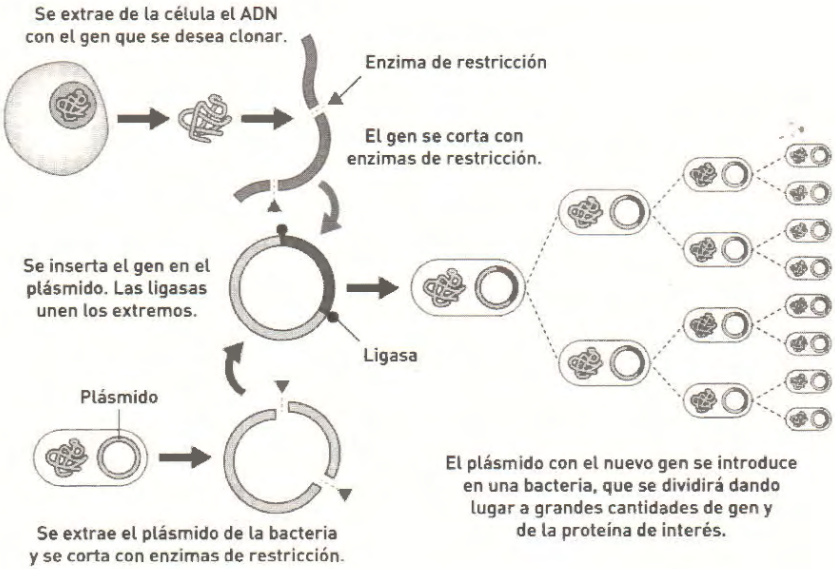


Figura que reproduce el mecanismo de producción de ADN recombinante: mediante el aislamiento del gen de interés es posible producir grandes cantidades de proteína.

tes diabéticos cada día. A comienzos del siglo XX, Frederick Grant Banting y Charles Best, dos científicos de la Universidad de Toronto (Canadá), descubrieron que la insulina era la hormona necesaria para tratar a pacientes diabéticos, por lo que comenzó a suministrarse insulina porcina, previamente purificada, para salvar la vida a estos pacientes; sin embargo, esta proteína causaba fuertes efectos adversos cuando era suministrada a humanos, ya que al ser una proteína exógena, contra la que el sistema inmunitario podía formar anticuerpos, algunos pacientes desarrollaron fuertes alergias. Frederick Sanger, en 1958, obtuvo la secuencia de los 51 aminoácidos de la insulina, tras más de una decena de años estudiándola, lo que permitió dejar de utilizar la insulina animal. De la secuencia de

aminoácidos, y siguiendo el código genético de manera inversa (de aminoácido a triplete) podría saberse con exactitud su secuencia genética. Hoy, la insulina humana se produce dentro de células de *Escherichia coli* en grandes tanques de cultivo.

Este no es el único ejemplo en el cual utilizamos fragmentos de ADN humano para introducirlos en microorganismos y que generen la proteína que codifican. Es también el caso de la hormona de crecimiento humana (HGH), actualmente utilizada en casos de hipo y acondroplasia (llamada vulgarmente «enanismo») y para aumentar la tasa metabólica en atletas. Esta se produce también en células de *Escherichia coli* y posteriormente es purificada para su administración a pacientes cuya glándula pituitaria, situada en la base del cerebro, genera cantidades insuficientes. Antes de que la hormona recombinante estuviera disponible, la hormona para uso terapéutico era obtenida de las glándulas pituitarias de cadáveres. De hecho, fue en 1956 cuando el endocrinólogo Maurice Raben aisló la HGH de cadáveres humanos y de monos; pero su acción terapéutica no fue la esperada, ya que muchos de los pacientes a los que se les suministró desarrollaron la enfermedad de Creutzfeldt-Jakob, un trastorno cerebral degenerativo. El descubrimiento del código genético hizo que se pudiera mapear esta proteína en nuestro genoma, copiarla y producirla a gran escala. La ingeniería genética hizo posible utilizarla terapéuticamente sin complicaciones.

Otro importante avance de la medicina, generado tras conocer más en profundidad el código genético, fue la creación de la vacuna contra la hepatitis B recombinante. La vacuna se basa en la inoculación de un pequeño trozo de la superficie del virus que provoca la reacción del sistema inmunitario, el cual genera anticuerpos que se establecen en la circulación sanguínea. Por lo general, los virus atenuados que se inoculan como vacunas se consiguen del cultivo *in vitro* del propio virus patógeno. Se da la circunstancia de que el virus de la hepatitis B es un poliovirus del ARN que no se puede cultivar *in vitro*. Por ello, fue necesario obtener los genes del virus

de la hepatitis B para producir, a partir de ellos, las proteínas víricas sin necesidad de cultivar el virus en sí. El aislamiento de uno de los genes víricos que codifican para la cápside y su posterior clonación en un plásmido fue uno de los hitos en la historia de la medicina.

El hecho de conocer cómo los organismos vivos expresamos y damos sentido a la molécula de ADN ha permitido a la ciencia ir un paso más allá para buscar soluciones a los fallos de codificación que los seres humanos podemos presentar y que desembocan en enfermedades en ocasiones mortales.

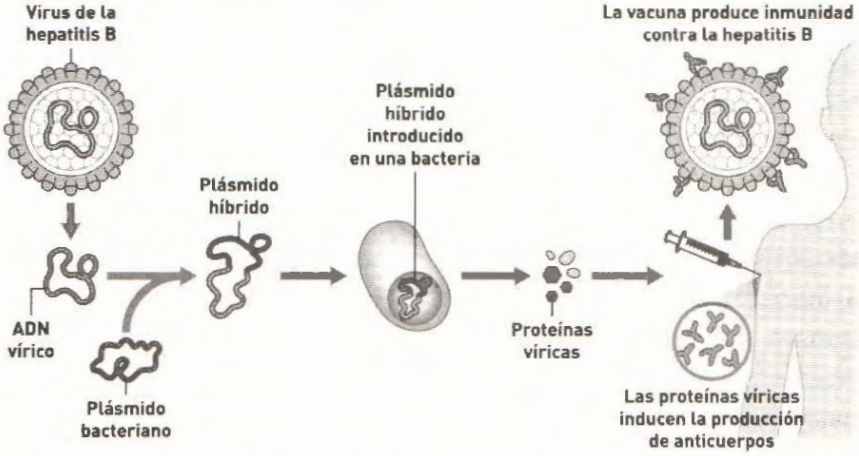
LA TERAPIA GÉNICA

Además de permitirnos leer y entender los fallos que pueden presentar nuestros genes, conocer el lenguaje de la molécula de la vida, el código genético, también nos ha llevado a ser capaces de reparar aquellas secuencias del mismo que contienen errores. Si esta reparación se lleva a cabo en las células de un organismo vivo para revertir algún fallo en el ADN, se denomina *terapia génica*. En la actualidad, estamos asistiendo a las primeras fases del desarrollo de este tipo de terapia que modifica lo más íntimo de nuestras células, los propios genes, afectando directamente al lenguaje del ADN, revertiendo o corrigiendo errores para sanar enfermedades.

Son ya muchos los ensayos clínicos que, modificando la información del ADN mediante las reglas que rigen nuestro código genético, han tenido resultados esperanzadores. Alteraciones como la diabetes, sida, insuficiencia cardíaca, alteraciones hepáticas, distrofias, arteriesclerosis y enfermedades neurodegenerativas como el alzhéimer presentan un horizonte de curación cada vez más próximo. La terapia génica clásica requiere una transferencia de genes sanos a células enfermas, de manera que los genes introducidos sean expresados en la cantidad adecuada. Estos genes, por lo general, se insertan en los cromosomas y se expresan como

> UNA VACUNA PARA LA HEPATITIS B

Las vacunas se basan en la capacidad de «memorizar» que tiene nuestro sistema inmune. De lo que se trata es de cultivar *in vitro* microorganismos patógenos atenuados o fragmentos del mismo para luego inocularlos en el organismo a fin de que este aprenda a reconocerlos como patógenos. El virus de la hepatitis B, sin embargo, no crece fuera de ningún organismo, por lo que no se podía manipular para generar vacunas. La solución llegó con el ADN recombinante, una técnica que permite transferir genes del virus de esa enfermedad (genes que codifican para antígenos y, por lo tanto, producen respuesta inmune) a organismos de fácil crecimiento como bacterias. De ahí, de la combinación de genes de dos organismos diferentes, deriva el nombre de «recombinante». Una vez se ha introducido el gen de la hepatitis B en una bacteria, esta comenzará a expresar las proteínas. En la actualidad, las bacterias se cultivan en grandes tanques que permiten la producción a escala industrial del antígeno de la hepatitis B, el mismo que una vez purificado se inocula en el organismo.



— Esquema del proceso de fabricación de una vacuna recombinante contra la hepatitis B: tras clonar una región del ADN del patógeno, esta se une al de una bacteria para crear un híbrido. Las proteínas resultantes, una vez purificadas, son usadas como vacuna.

uno más del genoma. Esto se ha empezado a llevar a la práctica en ensayos para tratar la enfermedad de alzhéimer, la cual afecta a más de cincuenta millones de personas en el mundo. Las líneas de trabajo emprendidas en los laboratorios de biomedicina molecular comienzan a dar buenos resultados en este campo. Uno de esos estudios, llevado a cabo por científicos del Imperial College de Londres, ha conseguido distribuir el gen PGC1-α en las células cerebrales de modelos de laboratorio con síntomas de alzhéimer. Al expresarse, PGC1-α aumenta la detoxificación de radicales libres de oxígeno en el cerebro de estos pacientes, retrasando la aparición de los síntomas. De hecho, los ratones tratados en las primeras etapas de la patología evolucionaron de manera muy favorable, siendo capaces de realizar tareas de memoria igual que los ratones sanos.

En esta terapia clásica, el gen insertado se integra en el genoma y se expresa en el tejido correspondiente. Esto confiere numerosas ventajas, ya que el gen así incluido puede perpetuarse por replicación en cada división celular, obteniéndose una expresión estable a largo plazo. En los tejidos formados por división activa de células, las investigaciones se centran en la capacidad de dirigir estos genes a las células madre, las cuales son las causantes de formar todas las células de un determinado tejido. Si logramos modificar a nivel genético las células madre, estaremos reprogramando en última instancia todas las células diferenciadas del tejido.

No obstante, la integración cromosómica tiene también importantes inconvenientes debido a que hasta ahora la inserción suele ocurrir casi al azar, es decir, el gen insertado puede hacerlo en diversas regiones del genoma. En algunos casos, los genes insertados pueden no expresarse debido a su inserción en regiones silenciadas; pueden también insertarse en regiones del genoma donde rompan la correcta lectura de otros genes, provocando en estos casos la muerte de la célula o, más grave aún, la aparición de enfermedades como el cáncer.

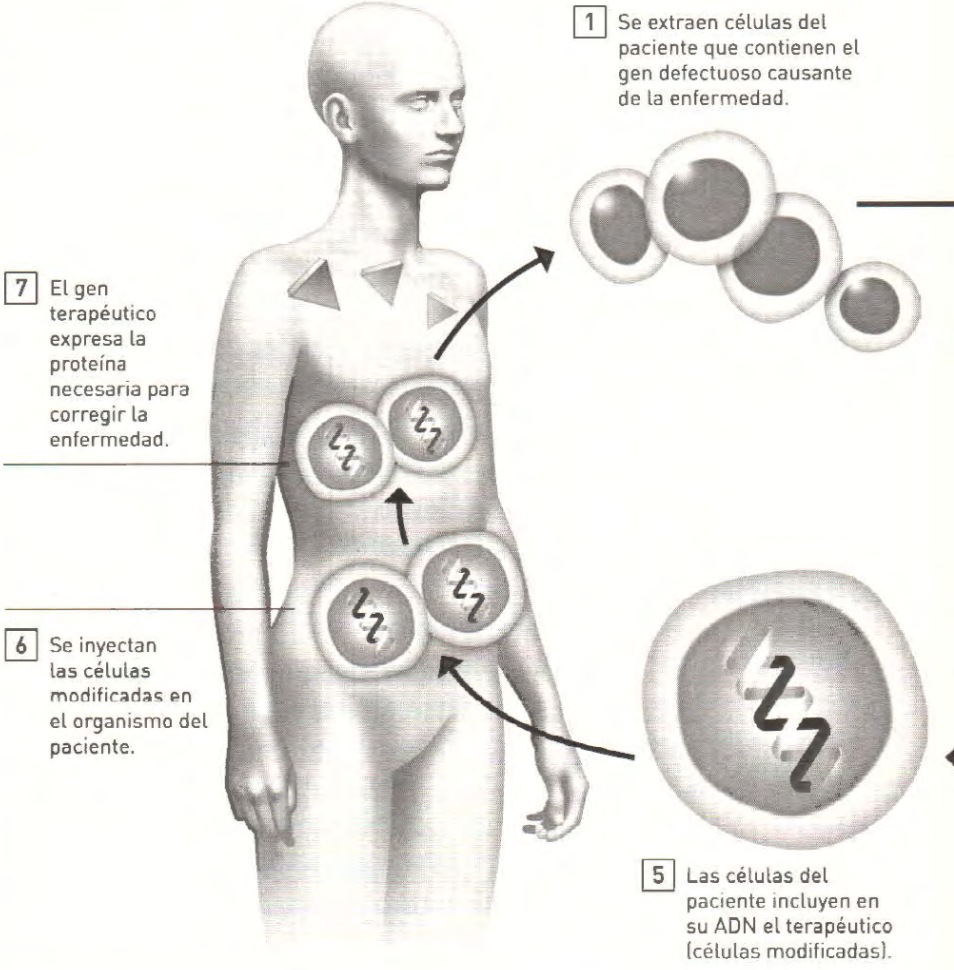
Para evitar esta situación, los laboratorios de medicina molecular han optado por la terapia génica *ex vivo*, lo que ofrece la oportunidad de seleccionar las células en las que la integración ha tenido éxito antes de su readministración al paciente.

De hecho, los primeros ensayos de terapia génica llevados a cabo con éxito se realizaron con células cultivadas *ex vivo* para poder seleccionar las células en las que la inserción había sido la correcta. Los pacientes fueron niños con inmunodeficiencia combinada grave (conocida por su sigla en inglés, SCID). En esta patología, las dos copias de la enzima adenosina desaminasa (ADA) están mutadas, de modo que producen proteínas no funcionales, lo que provoca la acumulación de desoxiadenosina. A determinadas concentraciones es tóxica, provocando la muerte de los linfocitos T del sistema inmune, lo que genera una inmunodeficiencia severa que deja al organismo expuesto ante todo tipo de agentes infecciosos, por lo que en numerosas ocasiones nos referimos a estos pacientes como «niños burbuja», ya que deben mantenerse aislados del exterior. Se han identificado más de 70 mutaciones en ADA que dan lugar a los síntomas de la enfermedad, la mayoría de las cuales son mutaciones de cambio de aminoácido (*miss-sense*) situadas en los exones 2 y 7 del gen. En este caso, un nucleótido cambia por otro, por lo que a la hora de interpretarse según las pautas de nuestro código genético, se genera un cambio de aminoácido que derivará en una proteína aberrante. Un ejemplo lo encontramos en la mutación Ala329Val, notación que indica que en la posición 329 de la proteína ADA existe una sustitución del aminoácido alanina por valina.

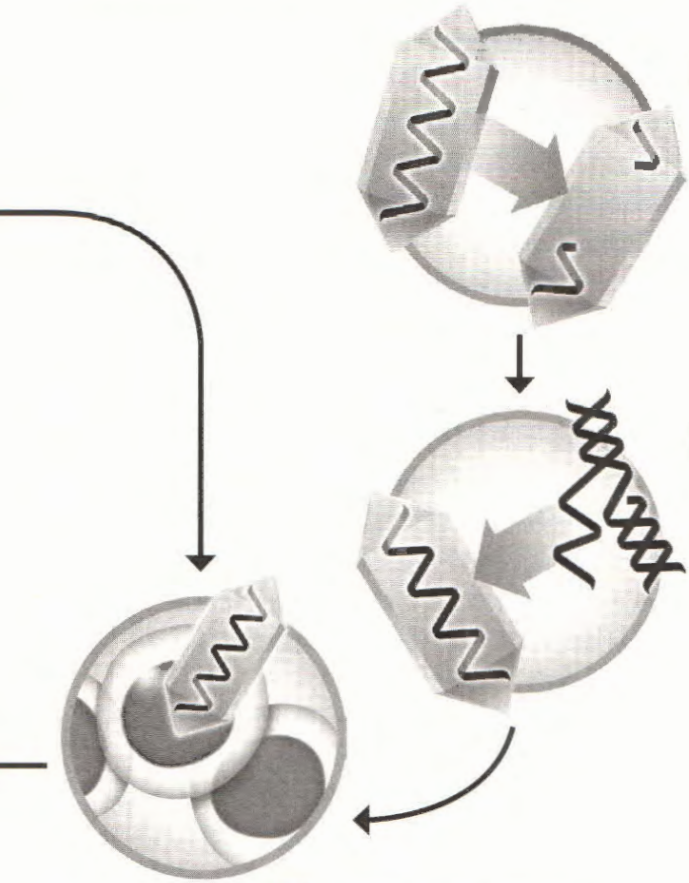
Vemos en este ejemplo que un simple cambio en un nucleótido puede traer consecuencias fatales para el organismo. La terapia genética clásica reintroduce el gen sano, pero es incapaz de corregir la mutación original. Por ello, la transferencia de genes tiene importantes limitaciones y no es aplicable en cualquier tipo de enfermedad genética. Se debe tener en cuenta que la naturaleza de la mutación causante de la enfermedad es de pérdida de función,

> CURAR LOS GENES UTILIZANDO VIRUS COMO VECTORES

El desciframiento del código genético ha abierto una nueva puerta a la biomedicina que hasta no hace tanto parecía impensable: reparar o incluso sustituir las secuencias de genes defectuosas que provocan las enfer-



medades de origen genético, y ello en las propias células del organismo afectado, como se observa en la figura. Es lo que se conoce como terapia génica. Aunque todavía no se ha desarrollado en todo su potencial, ya se ha demostrado su eficacia en muchos casos.



2 El material genético del virus, que se usa como vector, es tratado en el laboratorio para impedir que pueda reproducirse.

3 El gen terapéutico que nos interesa insertar en las células del paciente se encapsula en el virus.

4 El virus modificado con el gen terapéutico se introduce en las células extraídas del paciente.

como en el caso de la SCID, donde el defecto podría ser corregido por la introducción del gen no mutado para restablecer la función de la proteína si esta se expresa correctamente.

Por el contrario, las mutaciones de ganancia de función, como la aparición de una nueva proteína mutante o de un producto tóxico, que son características de algunas dolencias de tipo dominante, no pueden ser tratadas añadiendo genes normales y necesitarían de otras estrategias más difíciles de llevar a cabo, como el bloqueo específico del gen mutado o la corrección dirigida de la mutación.

En la investigación por conseguir una terapia genética eficaz, ha habido varios intentos para bloquear la función de genes activos de manera aberrante. Un ejemplo de ello lo encontramos en la acondroplastia, un tipo de enanismo causado por la actividad excesiva del gen FGFR3. Este gen gobierna la traducción de una proteína implicada en la conversión del cartílago en hueso durante el crecimiento. Si esta proteína actúa de manera muy activa, el cartílago se convierte rápidamente en hueso durante el período de crecimiento, impidiendo el desarrollo óptimo del individuo. El mal funcionamiento, en exceso, del gen FGFR3 se debe a una mutación puntual, que la mayoría de las veces se localiza en la posición del codón 380. Las primeras aproximaciones terapéuticas se basaban en dar fármacos inhibidores de la proteína FGFR3, pero la verdadera terapia de esta y muchas otras enfermedades tendrán que pasar por la eliminación de la mutación en la molécula de ADN, volviendo a establecerse así las instrucciones genéticas correctas.

Pero ¿cómo conseguir cambiar de manera dirigida un único nucleótido en el océano que representan los tres mil doscientos millones de nucleótidos de nuestro genoma? Esta visión de «buscar la aguja en un pajar» es hoy posible gracias a una herramienta que sin duda cambiará el futuro de la terapia genética y quizá la de la propia evolución humana. Hablamos de la ya citada técnica de edición genética CRISPR/Cas9, cuya denominación metafórica ha sido la de «tijeras moleculares con GPS» capaces de localizar con infinita

precisión la secuencia no deseada y eliminarla o sustituirla por otra en buen estado.

La revolución CRISPR

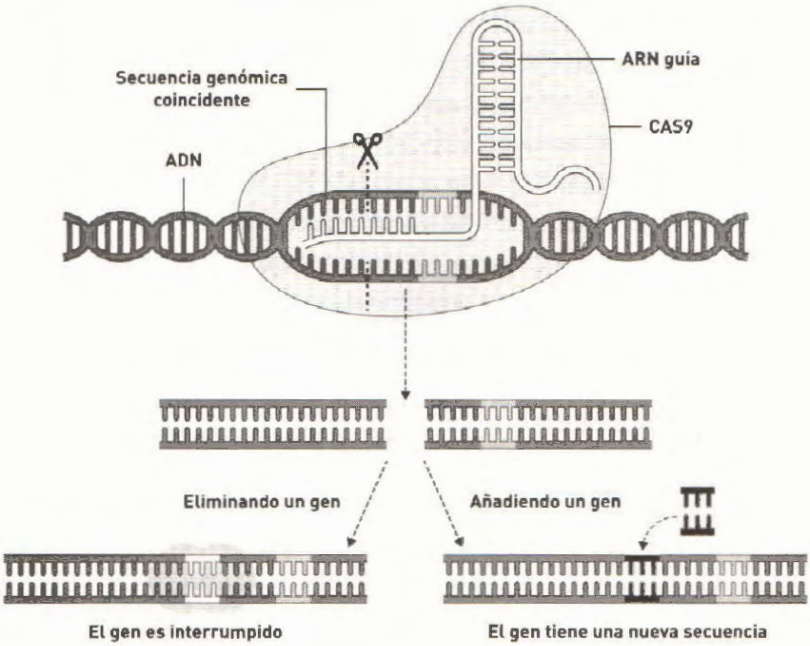
El descubrimiento de CRISPR/Cas9 ha sido uno de los más brillantes hallazgos que la ciencia básica nos ha brindado. A principios de la década de 1990, Francisco Juan Martínez Mojica, investigador en la Universidad de Valencia encontró unas regiones del genoma de las bacterias de la especie *Haloferax mediterranei* que se repetían a una distancia regular una de otra. Decidió seguir investigando en algo que, por aquel entonces, nadie apostaba por investigar. Lo que allí encontró fue un sistema de defensa bacteriano contra los virus capaces de atacarlas. Los virus, millones de veces más pequeños que las bacterias, se unen a la superficie de estas e inyectan su material genético directamente en su interior, utilizándolas como una fábrica de proteínas. En ocasiones, las bacterias sobreviven al ataque gracias a unas regiones repetidas en su genoma llamadas CRISPR (del inglés *clustered regularly interspaced short palindromic repeats*, «repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas»), capaces de almacenar parte del material genético del virus. Con ello, las bacterias desarrollan una memoria contra esos virus mediante la inserción del material genético vírico en su propio genoma. La próxima vez que la bacteria se exponga al virus, lo recordará y se defenderá mucho más rápida y eficazmente, ya que el ADN vírico guardado generará guías de ARN que podrán unirse al genoma del virus y combatirlo. Este proceso activa una enzima llamada Cas9, una endonucleasa capaz de (literalmente) cortar el ADN vírico.

> El CRISPR es una potente herramienta de edición genómica que puede ser programada para reconocer cualquier fragmento de ADN.
>
> EMMANUELLE CHARPENTIER

La verdadera revolución tuvo lugar cuando se comprendió que Cas9 se puede programar para generar un corte guiado en el ADN. En agosto de 2012, la química estadounidense Jennifer Doudna y la microbióloga y bioquímica francesa Emmanuelle Charpentier publicaron un artículo en la revista *Science* en el que demostraban que si generamos un ARN que luego unimos a Cas9, esta enzima buscará en el ADN la secuencia complementaria para cortarla. Como el ARN se genera en laboratorios de manera que sea complementario a la secuencia de ADN que queramos modificar podremos dirigir estas tijeras a voluntad. En realidad, lo que hacemos es sencillamente cambiar el material genético del virus por otro que nos sea de utilidad. Como consecuencia, Cas9 con el ARN sintetizado se dirige a la zona de ADN que nos interesa generando un corte de doble cadena. Este corte es unido de nuevo por la célula, utilizando uno de los mecanismos de reparación de los que ya nos hemos referido con anterioridad. Además, este método también puede servir para introducir un fragmento de ADN de interés en una región específica. Para ello debemos introducir en las células el fragmento que queramos insertar, incluyendo en el ARN que se une a Cas9 la secuencia adecuada para que inserte exactamente las bases complementarias para que en el ADN de la célula se produzca la proteína deseada (fig. 2).

El hecho de conseguir corregir una mutación directamente en el ADN sin la necesidad de administrar genes exógenos elimina también el problema de la regulación de la expresión. Debemos tener en cuenta que no todo reside en la lectura de los genes, en su descodificación a proteína, sino que los genes suelen requerir un fino control de su expresión, necesitando niveles muy regulados de producto génico. Este es el caso de la β-talasemia, una enfermedad de los glóbulos rojos en los que no se sintetiza correctamente una de las cadenas que componen la hemoglobina, la β-globina. La hemoglobina se compone por dos cadenas, la α-globina y la β-globina, las cuales se unen en proporción 1:1 para formar una hemoglobina funcional. Con niveles más bajos de β-globina, la hemoglobina resultante no

Fig. 2



La proteína Cas9, con ayuda de un ARN guía, localiza dentro del genoma la región de interés. Tras localizarlo tiene la capacidad de cortar la secuencia generando bien una deleción o bien la introducción de un nuevo fragmento de ADN.

libera oxígeno de forma eficiente en los tejidos, presentándose problemas de hipoxia. La expresión del gen de la β-globina insertado requiere, para curar la enfermedad, un control exacto de su producción, que debe ser igual al de la β-globina, lo que precisa un fino control de la expresión génica que ha impedido hasta el momento la consecución de progresos significativos en la aplicación de la terapia génica a esta enfermedad sanguínea. También aquí, la llegada de CRISPR ha revolucionado el modo que tenemos de entender la terapia genética, pasando de suministrar genes a modificar, a sanar los que ya tenemos. En la actualidad, ya existen varios ensayos clí-

nicos con CRISPR para recuperar la forma original de la β-globina en los glóbulos rojos mediante la manipulación de células madre hematopoyéticas con CRISPR, lo que elimina el problema de la precisión a la hora de regular la cantidad, ya que de este asunto se encargan las secuencias genómicas adyacentes al gen, los promotores.

La irrupción de CRISPR ha supuesto un impresionante avance para la biomedicina molecular, ya que permite editar el ADN de las células de cualquier organismo, incluidos los humanos. Las primeras pruebas efectuadas con animales apuntan a que esta técnica pronto será utilizada de manera importante a nivel clínico, abriendo al mundo la capacidad para corregir defectos genéticos producto de mutaciones causantes de enfermedades muy graves.

Hacia el final de las enfermedades genéticas

Poco tiempo después de descubrir la posibilidad de programar CRISPR/Cas9 en los laboratorios, los científicos demostraron su potencial en terapia genética tras cortar y deshacerse del ADN del virus del sida, el VIH, que infecta los linfocitos T en humanos. El experimento fue realizado en células en cultivo y pronto pasó a ser realizado con animales de laboratorio, en los que se consiguió inhibir la replicación del virus gracias a un corte estratégico en su material genético. Tan solo un par de semanas después de la terapia, el nivel de virus bajó drásticamente en los animales tratados. Actualmente ya se están desarrollando potentes antivíricos basados en la técnica CRISPR/Cas9 que prometen corregir problemas como la fibrosis quística, la hepatitis B, el herpes o el sida en un futuro nada lejano.

Las aplicaciones en las que podemos pensar son innumerables. De hecho, en 2016 se publicaron simultáneamente tres trabajos en los que se decía que CRISPR/Cas9 podría recuperar parcialmente la capacidad de contracción de los músculos perdida en la distrofia muscular de Duchenne. Esta enfermedad se produce por mutacio-

> LAS GRANDES IMPULSORAS DEL CRISPR

La francesa Emmanuelle Charpentier, directora de un laboratorio de investigación en la Universidad de Umea (Suecia) y en el Centro Helmholtz para la Investigación de Infecciones (Alemania), y la estadounidense Jennifer Doudna, profesora en la Universidad de Californa en Berkeley y directora de su propio laboratorio en el Howard Hughes Medical Institute (EE.UU.), publicaron en 2012 un artículo en el que describían una técnica, el CRISPR/Cas9, para editar genes que iba a revolucionar el campo de la biomedicina. En realidad, su descubridor fue el científico español Francisco Juan Martínez Mojica, pero fueron ellas quienes consiguieron programar en el laboratorio los lugares precisos de corte de Cas9, es decir, en qué lugar exacto del genoma añadir o quitar material genético. En la actualidad, Doudna y Charpentier trabajan con el apoyo de importantes inversores y laboratorios para desarrollar futuras aplicaciones médicas de esa técnica.

Emmanuelle Charpentier (a la izquierda) y Jennifer Doudna recibieron en 2017 el premio Fronteras de Conocimiento de la Fundación BBVA.

nes en el gen de la distrofina, uno de los genes más grandes conocidos, con 79 exones. La mayoría de las mutaciones que afectan al gen provocan un cambio en el patrón de lectura, generando una proteína aberrante con un codón de stop prematuro. Los ensayos con CRISPR en los que se restablece parte de la funcionalidad de la distrofina se basan en la eliminación completa del exón que porta la mutación. Así, el ARN resultante carecerá de esa región exónica, pero conservará la pauta de lectura correcta en el resto de la proteína. Otras estrategias han conseguido reparar la mutación por completo, mediante el uso de una plantilla de ADN a partir de la cual corregir. De momento, estos procedimientos únicamente se han realizado con animales de laboratorio, aunque es solo cuestión de tiempo para que su aplicación se realice en humanos.

La tecnología CRISPR/Cas9, no obstante, ya ha llegado a la terapia con humanos: en agosto de 2016 unos laboratorios de China llevaron a cabo el primer tratamiento de este tipo con pacientes de cáncer de pulmón. El experimento se basaba en potenciar el sistema inmune. Para ello se extrajeron células inmunes del paciente para cultivarlas *ex vivo* (que, como hemos visto, da mejores resultados). En ellas se realizó un corte dirigido en su genoma, en todas el mismo, que eliminaba la expresión de la proteína PD-1, la cual es capaz de frenar el ataque del sistema inmune a agentes externos. Eliminando este «punto de freno» las células son capaces de atacar al tumor de manera mucho más eficiente, tras su inyección nuevamente en el paciente. Esta terapia consiguió reducir eficientemente el tamaño del tumor.

La lista de enfermedades genéticas que podríamos llegar a tratar con CRISPR es enorme: hemofilia, Huntington, fibrosis quística, enfermedades cardiovasculares, diabetes tipo I e incluso daltonismo, y todo porque existen muchas enfermedades causadas por un único cambio en la secuencia de nucleótidos del ADN. En esa molécula con tres mil doscientos millones de letras, un pequeño cambio de una G por una A, o de una T por una C (o cualquier otra

posibilidad), una deleción o pequeñas inserciones puede llevarnos a una enfermedad en ocasiones mortal. Es por ello que la posibilidad de editar nuestros genes para mantener una lectura óptima se presenta como una prioridad para asegurar un futuro libre de enfermedades genéticas.

El envío de CRISPR al interior celular

El sistema CRISPR/Cas9 ha revolucionado a todas luces la investigación biomédica. Sin embargo, es un sistema bacteriano situado en el genoma de estos organismos procariotas, por lo que uno de los mayores obstáculos que se está encontrando para su correcto desarrollo es el método de transferencia a células humanas. Nuestras células no cuentan con ningún tipo de proteasa similar a Cas9, por lo que de alguna manera debemos transferirlas al núcleo de estas.

Uno de los sistemas más utilizados hasta la fecha para transferir genes a células en un organismo vivo es el uso de virus, concretamente adenovirus, como vectores y, en concreto, de los adenovirus (fig. 3).

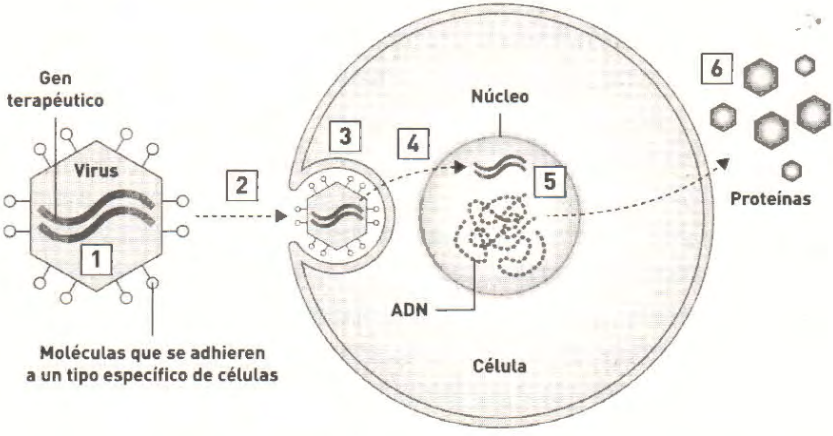
Estos virus, en condiciones naturales, son los causantes de patologías frecuentes como el resfriado común. Una de las ventajas de encapsular genes dentro de adenovirus para que los integren en las células diana es que pueden transferir grandes porciones de ADN a casi cualquier célula. Además, estos genes no se integran en el genoma de la célula hospedadora, por lo que no hay peligro de que rompan la pauta de lectura de algún gen esencial para el correcto desarrollo celular. Esto puede suponer una desventaja, ya que al no integrarse el gen de interés en los cromosomas celulares, se pierde en la primera división celular. Además, este tipo de virus genera una fuerte reacción inmunológica en los pacientes, por lo que su uso parece cada vez más problemático. Las técnicas futuras apuntan a un cambio radical en los vehículos de transferencia de genes.

Fig. 3

1 Se inserta un gen terapéutico en un vector viral.

2 El virus se inocula al paciente.

3 El virus afecta a la célula diana.



4 Una vez dentro del núcleo, el virus descarga el gen.

5 El gen se une al ADN de la célula.

6 La célula produce nuevas proteínas que mejoran los síntomas de la enfermedad.

Los adenovirus son capaces de transferir ADN a las células el paciente. Este ADN podrá generar las proteínas de las que las células del paciente carecen.

En este sentido, ya hay numerosos equipos de investigación mirando al horizonte futuro y desarrollando nuevas técnicas de transferencia de genes. Uno de los ensayos más exitosos es el que ha utilizado nanopartículas lipídicas para transferir los componentes del sistema CRISPR/Cas9 de una manera muy eficaz a células del hígado. Un equipo de investigadores, asentado en China, ha logrado englobar en nanopartículas de grasa el ARNm de Cas9 y el ARN guía capaz de dirigir la proteasa Cas9 al gen PCSK9, un gen alterado en la hipercolesterolemia familiar. Las pequeñas nanopartículas logran transferir muy eficazmente estos ARN al inte-

rior de las células hepáticas, donde se expresa PCSK9, una proteína implicada en la captación de LDL (*low density lipoprotein*) por parte de las células. El LDL, denominado colesterol malo, es el que se encuentra aumentado en la hipercolesterolemia familiar. En algunas familias, el problema de esta enfermedad reside en mutaciones de ganancia de función de PCSK9, por lo que reduciendo la cantidad de esta proteína disminuirá el nivel de colesterol. Es por ello que el ARN guía está específicamente diseñado para cortar PCSK9, provocando deleciones que alteran la pauta de lectura a partir del corte.

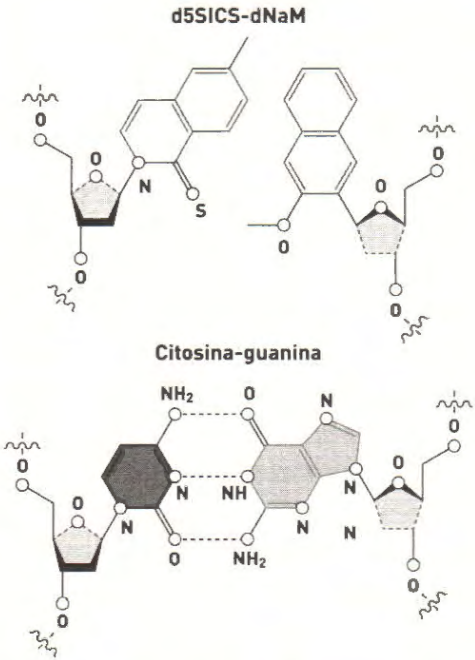
No es descabellado pensar que el futuro de la biomedicina molecular pasa por la utilización de CRISPR/Cas9, una tecnología impensable sin un profundo conocimiento de nuestro ADN y las leyes y códigos que lo gobiernan.

Sin embargo, en ocasiones la investigación no solo quiere comprender las actuales leyes, sino que busca también cambiar las reglas del juego. Una vez que hemos comprendido cómo funciona el código que gobierna nuestro genoma, la ciencia ahora reta a la naturaleza añadiendo caracteres para formular un código distinto, incrementando así su complejidad. Quizás estemos en los albores de construir nuevas leyes que gobiernen la vida.

RETANDO A LA NATURALEZA: UN CÓDIGO GENÉTICO ARTIFICIAL

Toda la vida que conocemos sobre este planeta se gobierna por cuatro letras (tripletes de nucleótidos). La evolución ha seleccionado de manera apabullante este tipo de código vital en detrimento de otros que hayan podido surgir, evidentemente, sin tanto éxito. Estamos en un momento en el que conocemos el lenguaje, sus lecturas y su mecanismo de edición. Pero ¿qué pasa si modificamos el código, si introducimos nucleótidos artificiales? Esto ya se ha

Fig. 4



Arriba, pareja de nucleótidos sintéticos formada por bases hidrofóbicas que contrastan con la pareja citosina-guanina formada por enlaces de hidrógeno.

realizado en laboratorios. A los cuatro nucleótidos conocidos se les han añadido dos más, artificiales, formados por bases hidrófobas, ejemplificadas por el par formado entre d5SICS y dNaM (d5SICS-dNaM), que pueden incorporarse a la replicación y ser transcritos *in vitro* (fig. 4).

El responsable de esta modificación fue Floyd E. Romesberg, del Instituto de Investigación Scripps de California, quien quiso demostrar así que es posible expandir la codificación de la vida más allá del punto al que llegó la evolución. De ese modo logró una bacteria que podía incluir sin mayor problema los nucleótidos no naturales en su ADN y copiarlos generación tras generación. Este es solo un ejemplo

de la potencialidad de conocer y expandir el alfabeto de la vida, ya que nos permitiría crear nuevas formas vivas cuyo ADN se escriba y programe como el sistema operativo de un ordenador. Pero la idea va más allá: no solo podemos integrar nuevos nucleótidos en el ADN, sino llegar también a fabricar proteínas que tengan en su estructura nuevos componentes que no se encuentran en la naturaleza.

En definitiva, el conocimiento del código de lectura del ADN, la generación de herramientas para copiarlo y editarlo, así como la capacidad de introducir nuevas letras en el lenguaje celular son elementos que hacen que el futuro de la biomedicina tenga que pasar por la aplicación de las recientes técnicas de la genética molecular, entrando en un ámbito lleno de retos en el que el conocimiento y la imaginación podrán modificar y reprogramar la expresión genética de las células, lo que sin duda nos permitirá liberarnos de muchas enfermedades que actualmente no tienen solución, y quizá también diseñar a voluntad o, al menos moldear, nuestra información genética.

LECTURAS RECOMENDADAS

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., *Biología molecular de la célula*, Barcelona, Omega, 2010.

Esteller, Manel, *No soy mi ADN*, Barcelona, RBA, 2017.

Griffiths, A. J. F., Wessler, S. R., Suzuki, D. T., Lewontin, R. C., Carroll, S. B., *Genética*, Madrid, McGraw-Hill/ Interamericana de España, 2008.

Henderson, Mark, *50 cosas que hay que saber sobre genética*, Barcelona, Ariel, 2016.

Herráez, Ángel, *Texto ilustrado de biología molecular e ingeniería genética*, Barcelona, Elsevier, 2012.

Jones, S., Van Loon, B., *Genética para todos*, Barcelona, Paidós, 2005.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Ploegh, H., Amon, A., Scott, M. P., *Biología celular y molecular*, Madrid, Médica Panamericana, 2016.

Mukjerjee, Siddhartha, *El gen una historia personal*, Barcelona, Debate, 2017.

Pita, Miguel, *El ADN dictador: lo que la genética decide por ti*, Barcelona, Ariel, 2017.

Smith, Gina, *Qué somos realmente*, Barcelona, Robin Book, 2005.

Soberon Mainero, Francisco Xavier, *La ingeniería genética, la nueva biotecnología y la era genómica*, Madrid, Fondo de Cultura Económica de España, 2012.

Strathern, Paul Crick, *Watson y el ADN*, Madrid, Siglo XXI, 2014.

Stryer, L., Berg, J. M., Tymoczko, J. L., *Bioquímica: con aplicaciones clínicas*, Barcelona, Reverté, 2013.

Valpuesta, José María, *A la búsqueda del secreto de la vida*, Madrid, Hélice, 2008.

Watson, J., Berry, A., *ADN: el secreto de la vida*, Madrid, Taurus, 2003.

ÍNDICE

adenina 8, 18-19, 23-24, 28, 46, 89
adenovirus 94, 129-130
ADN ligasa 88
ADN polimerasa 87-89, 98-99
agentes genéticos 93
anemia falciforme 20-21
anticodón 29, 30-31, 62, 64
apoptosis 107
arginina 33, 36-38, 60, 80, 110
ARN polimerasa 23-24
ARNm (ARN mensajero) 9-10, 22-33, 36, 51, 56, 58, 61, 69-71, 79-80, 130
ARNt (ARN de transferencia) 28-31, 38-39, 53, 62, 64, 69-72
Astrachan, Lawrence 54
Avery, Oscar Theodore 45-46

Brenner, Sydney 52, 54, 56

cadenas poli(U) 59
capping 26
célula
  eucariota 27
  germinal 78
  reproductora 78, 85
Chargaff, Erwin 47
citoplasma 24-26, 28, 30-32, 47, 55
citosina 8, 18-19, 23-24, 46, 64, 89, 92, 132
cloroplasto 38
Club de las Corbatas de ARN 9, 47-49, 60
Cockcroft, John 49
codón de inicio 34, 36

codón stop 31, 33, 37-39, 68, 80, 82, 110, 128
cola de poli-A 28
cromátidas 86

daltonismo 81, 128
Delbrück, Max 47
deleciones 60, 82-83, 85, 88, 100, 108, 110, 131
desaminación 89, 92
distrofina 82, 128
doble cadena 19, 22, 25, 32, 46, 87, 93, 124

empalme alternativo 26-27
estrés oxidativo 92

factores de liberación 31
Feynmann, Richard 47
fragmentos de Okazaki 87-88

Gamow, George 9, 47-49, 62
glicina 36, 39-40, 64, 83
glutamina 37, 64, 82

helicasa 87-88
Khorana, Har Gobind 63, 69
Holley, Robert 62-63, 69

intrones 24, 26
isoleucina 36, 38, 40, 60

Jacob, François 54, 56

MacLeod, Colin 46
Margulis, Lynn 35
Matthaei, Heinrich 56, 58, 59, 63
McCarty, Maclyn 46
McClintock, Barbara 93, 95
meiosis 85-86
Mendel, Gregor 9
Meselson, Matthew 56
Morgan, Thomas Hunt 45
mutación
  de cambio de sentido 80
  genética 87
  silenciosa 80
  sin sentido 67, 80
  somática 79

Ochoa, Severo 56-58, 63-64
Orgel, Leslie E. 49

pauta de lectura 46, 48, 60-61, 82-83, 108, 128-129, 131
polinucleótido fosforilasa 56-57, 59, 64
Proyecto Genoma Humano 16, 72, 111
puentes de hidrógeno 19, 87

radiación ultravioleta 89-90, 97
rayos
  gamma 91-92
  X 91-92, 100
recombinación meiótica 85-86, 101

retrotranscripción 93-94, 96
retrotransposones 93
ribonucleótidos 9, 22
Rich, Alexander 49
Roberts, Richard J. 26

Sharp, Phillip Allen 26
síndrome de
  Cockayne 98
  del *cri du chat* 84
  Down 84
  Klinefelter 84
  Lynch 99, 108
  Turner 84
  Williams 84
síntesis proteica 22, 58

terapia genética 8, 119, 122, 125-126
timina 8, 18-19, 22-23, 46, 90, 99
tirosina 37, 39, 69
translocación 83, 85
transposones 93-94
treonina 36, 60, 64, 69
tricotiodistrofia 98

Volkin, Elliot 54

Watson, James 9, 46-47, 49

Zamecnik, Paul 50, 53